433

Frontiers in Artificial Intelligence and Applications



DIGITALIZATION AND MANAGEMENT INNOVATION V

Proceedings of DMI 2026, Xi'an, China, 24-27 August 2026

Edited by
Adela Lau
Antonio J. Tallón-Ballesteros

IOS Press

DIGITALIZATION AND MANAGEMENT INNOVATION V

Frontiers in Artificial Intelligence and Applications

The proceedings published in the *Frontiers in Artificial Intelligence and Applications* series (*FAIA*) cover a broad spectrum of artificial intelligence research, from theoretical foundations to applied and interdisciplinary applications. Across the series, three recurring areas stand out. First, FAIA presents foundational AI research, including work on machine learning, reasoning, knowledge representation, logic, and intelligent algorithms, which underpins progress across the field. Second, it showcases applied AI and intelligent systems, where AI techniques are developed, tested, and deployed in domains such as robotics, data analytics, software engineering, healthcare, law, and engineering. Third, *FAIA* highlights interdisciplinary and human centred perspectives, exploring how AI interacts with humans, organisations, and society through themes such as explainable and trustworthy AI, human–AI interaction, hybrid intelligence, ethics, and responsible innovation. Drawing on contributions from leading international conferences, the series provides a comprehensive and evolving picture of contemporary artificial intelligence research and practice.

Series Editors:
Nicola Guarino, Pascal Hitzler, Joost N. Kok, Jiming Liu, Ramon López de Mántaras,
Riichiro Mizoguchi, Mark Musen, Sankar K. Pal, Ning Zhong

Volume 433

Recently published in this series

Vol. 432. L. Trajković, V.C.S. Lee and J. He (Eds.), Education, Network and Information Technology – Proceedings of the 9th International Conference (ICENIT 2026), Zhengzhou, China, 24-26 July 2026
Vol. 431. V. Costa, X. Angerri Torredeflot, J. Vázquez-Salceda, E. López-Iñesta, L. Museros and I. Sanz (Eds.), Artificial Intelligence Research and Development – Proceedings of the 28th International Conference of the Catalan Association for Artificial Intelligence
Vol. 430. H. Fujita, A. Selamat and F. Mohamed (Eds.), New Trends in Intelligent Software Methodologies, Tools and Techniques – Proceedings of the 25th International Conference on New Trends in Intelligent Software Methodologies, Tools and Techniques (SoMeT_26)
Vol. 429. J.P.A. Almeida, F. Baião, S. de Cesare and T.P. Sales (Eds.), Formal Ontology in Information Systems – Proceedings of the 16th International Conference (FOIS 2026), Vitória, Brazil, 21-25 September 2026
Vol. 428. K. Atkinson, M.V. Martinez and L. Godo (Eds.), Computational Models of Argument – Proceedings of COMMA 2026
Vol. 427. J. Ma (Ed.), International Symposium on Artificial Intelligence – Proceedings of the 6th International Symposium (ISAI 2026), Chengdu, China, 24-26 April 2026

ISSN 0922-6389 (print)
ISSN 1879-8314 (online)

Digitalization and Management Innovation V

Proceedings of DMI 2026, Xi'an, China, 24-27 August 2026

Edited by

Adela Lau

The University of Hong Kong, China

and

Antonio J. Tallón-Ballesteros

University of Huelva, Spain

Sage | IOS Press

© 2026 The Authors.

This book is published online with Open Access and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).

ISBN 978-1-64368-695-0 (online)
doi: 10.3233/FAIA433



IOS Press

A Sage company
Teleportboulevard 120
1043 EJ Amsterdam

Sage
3rd Floor, HYLO
103–105 Bunhill Row
London, EC1Y 8LZ

Sage
3011 Townsgate Road, Suite 120
Westlake Village, CA 91361

Sage
Unit No 323-333, Third Floor, F-Block
International Trade Tower
Nehru Place, New Delhi – 110 019

Sage
8 Marina View Suite 43-053
Asia Square Tower 1
Singapore 018960

DISCLAIMER
The authors, editors, and publisher will not accept any legal responsibility for any errors or omissions that may be made in this publication. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Preface

The 5th International Conference on Digitalization and Management Innovation (DMI 2026), which focuses on the general theme of *digitalization and management innovation* along with cutting-edge research topics, drew substantial participation from professors and postgraduate students—particularly PhD scholars—across global universities, as well as technicians, policymakers, and senior executives from diverse industries and organizational sectors.

This publication constitutes the proceedings volume, which is structured into five parts that reflect the conference's thematic breadth:

- Part I: Digitalization and Management Innovation (the foundational track aligned with the conference's core mission);
- Part II: Track on "Applied-education and applied-research in AI, Data Science, and Web6.0", the Second Edition (focused on bridging academia and practice in emerging tech domains);
- Part III: Special Session on "New Ecology of Lifelong Education" (exploring digitalization's role in reshaping lifelong learning systems);
- Part IV: Special Session on "The Potential of Big Data, Digitalization and AI in a Sustainable Economy" (addressing tech-driven solutions for sustainable development);
- Part V: Interdisciplinary Applications of Digitalization and Management Innovation (showcasing cross-field adaptations of digital management practices).

All submissions underwent rigorous and meticulous peer review by international Technical Program Committee (TPC) members and anonymous reviewers, ensuring alignment with DMI's research scope in terms of both thematic breadth and analytical depth. From 273 submissions, **83 high-impact contributions** with strong relevance to the *Frontiers in Artificial Intelligence and Applications* (FAIA) series were selected for inclusion, representing the most promising advancements in the field, including topics such as digital transformation in organizational/enterprise management; AI and machine learning-enabled management analytics; digital governance and innovation ecosystems; talent development for digital literacy and Smart city management innovation.

Finally, we extend our sincere gratitude to all keynote speakers, invited presenters, authors, TPC members, and anonymous reviewers, whose dedicated efforts were instrumental in the conference's success. Additionally, we would like to acknowledge the editorial team and colleagues at Sage Publishing / IOS Press for their collaboration and hard work in publishing this volume as the fifth edition of the DMI conference series within the FAIA series.

September 2026

Adela Lau, The University of Hong Kong, China
Antonio J. Tallón-Ballesteros, University of Huelva, Spain

About the Conference

Co-organized by Beijing Wuzi University, Xiangnan University, and HKU SAAS Data Science Lab, The University of Hong Kong, the 5th International Conference on Digitalization and Management Innovation (DMI 2026) took place from August 24 to 27, 2026, both on-site in Xi'an, China and online. DMI 2026 aimed to provide a high-level platform for experts and scholars worldwide to share their latest ideas on digitalization and management innovation, and featured a Special Session on "New Ecology of Lifelong Education" which was supported by Hunan Lifelong Education Research Base Xiangnan University Base, Special Session on "The Potential of Big Data, Digitalization and AI in a Sustainable Economy" (proposed and chaired by Prof. Yoshihiro Hamaguchi, Professor of Macroeconomics at Hannan University, Osaka, Japan), and a Track on "Applied-education and applied-research in AI, Data Science, and Web6.0", the Second Edition (proposed and chaired by Dr. Adela Lau from HKU SAAS Data Science Lab, School of Computing and Data Science, The University of Hong Kong) to foster the prosperity of the discipline. The agenda of the conference consisted of keynote speeches, invited speeches and oral presentations.

Conference Co-Chairs

Prof. Jin Chen**, Dean, Business School, Beijing Wuzi University, China**
Prof. Weiping Liu**, Secretary of CPC Committee, Xiangnan University, China**

Technical Program Committee (TPC)

Dr. Otmane Azeroual, Senior Researcher and Lecturer, German Centre for Higher Education Research and Science Studies (DZHW), Germany
Dr. Antonio J. Tallón-Ballesteros, Professor, Department of Electronic, Computer Systems and Automation Engineering, University of Huelva, Spain
Dr. Cristian Barra, Associate Professor of Economics, Department of Economics and Statistics (DISES), University of Salerno, Italy
Dr. Rocky J. Dwyer, Professor, College of Management and Human Potential, Walden University, USA
Dr. Ljupcho Efremov, American University of the Middle East, Kuwait
Dr. Goh Chin Fei, Senior Lecturer, Universiti Teknologi Malaysia, Malaysia
Dr. Galina Ilieva, Professor, Department of Management and Quantitative Methods in Economics, University of Plovdiv Paisii Hilendarski, Bulgaria
Dr. Adela Lau, Deputy Director of Data Science Lab, School of Computing and Data Science, The University of Hong Kong, China; Director of Dr. Adela Education Limited
Dr. Chien-Sing Lee, Professor, School of Computing and Artificial Intelligence, Sunway University, Malaysia
Dr. Jesús Barrena-Martínez, Associate Professor, Faculty of Business and Economics, University of Cadiz, Spain

Dr. Giuseppe Orlando, Professor, Department of Economics and Finance, Department of Mathematics, University of Bari, Italy
Dr. Víctor-Hugo Perera Rodríguez, Professor, Universidad de Sevilla, Spain
Dr. Angeles Sánchez, Associate Professor, Department of Applied Economics, University of Granada, Spain
Dr. João Rodrigues dos Santos, Professor, Universidade Europeia, Portugal
Dr. Gautam Srivastava, Professor, Department of Computer Science, Brandon University, Manitoba, Canada
Dr. Razia Sulthana, Senior Lecturer, School of Computing & Mathematical Sciences (CMS), Faculty of Engineering and Science (FES), University of Greenwich, Old Royal Naval College, UK
Dr. Jakkrit Thavorn, Assistant Professor, Faculty of Commerce and Accountancy (Thammasat Business School), Thammasat University, Thailand
Dr. Peter Trim, Reader in Marketing and Security Management, Birkbeck Business School, Birkbeck, University of London, UK

*Listed in alphabetical order by last name

Contents

Preface v
Adela Lau and Antonio J. Tallón-Ballesteros

About the Conference vi

Section 1. Digitalization and Management Innovation

An Evaluation Study on the "Five-Dimensional Integrated" Digital Intelligence E-Commerce Talent Cultivation Model Based on AHP-TOPSIS 2
Yan Wang

The Adoption and Impact of Smart Wearable Devices: A Literature Review Based on the "Antecedent-Mechanism-Consequence" Framework 15
Yawei Wang and Qinmei Wu

A Study on the Influencing Factors of Tacit Knowledge Sharing Intention Among Female Hotel Managers: A TPB-Based Perspective 24
Yidan Zhu, Yingjie Liu and Yuqian Liu

The Impact of Digital Transformation on Corporate Market Value: The Moderating Effect of Top Management Team Heterogeneity 35
Ziyang Liu, Shijie Bai and Jing Sun

A Study on Generation Z's Information Reception Behaviour Regarding Pre-Prepared Meals in an Algorithm-Driven Recommendation Context 48
Zengjingwen Chen, Yufang Peng, Yuchen Zhang, Bei Liu and Qi Liu

Revisiting Managerial Rationality: Structural Limitations of Agent-Based Artificial Intelligence in Executive Decision-Making 64
Mikhail Komarov and Aleksandr Korostelev

AI Governance Maturity and Enterprise Adoption of Agentic AI: Evidence from a Mixed-Methods Study 76
Sergei Emelianov

Labor Relations and Labor Rights Crisis Under Digital Supply Chain Disruption: A Case Study of Shein's Fast Fashion Digital Supply Chain 86
Haishui Jin and Jianxiu Wang

The Future of E-Business in the Era of the Technological Singularity 97
Dimiter Velev, Plamena Zlateva and Marin Galabov

Landscape Narrative Structure of Zhangjiajie Homestays: Evidence from UGC and Deep Learning 112
Yao Lu, Yidan Zhu and Yanan Liu

The Application of AI and Innovation Efficiency in Small and Medium-Sized Enterprises: A Perspective Based on Dynamic Capabilities and Resource Reallocation 123
Yichen Wu

Research on the Influencing Mechanism of Participation Intention in Immersive Red Tourism: A Case Study of the Guangdong–Hong Kong–Macao Greater Bay Area 130
Xiaoqi Huang, Hao Wu, Shuang Li, Weiqin Wu and Jingqi Lei

Digital Management Innovation for Back-End Publishing Operations in Engineering Construction: A Lightweight Transformation Framework Based on Company A 142
Jing Song

Demand Forecasting of Logistics Management Talents in Guangxi Under Digital-Intelligent Transformation 154
Yanni Liang, Xiaoqin Wang, Shanshan Zheng and Juan Huang

Research on the Application Practice Methods of Data Intellectual Property Protection 162
Qin Lei and Yajuan Tian

The Impact of Business-Trip Duration on Turnover Intention of Knowledge Workers: A Mediation Analysis 170
Yong Li and Guanghe Li

A Novel Supply Chain Packaging Design Evaluation Method via Sentiment Analysis of Reviews 180
Yangqing Liu, Haifeng Song, Xiumeng Li, Yaxi Ma, Jingqing Du, Shenghai Ke and Weiwei Yang

AI-Empowered Value Co-Creation for Intangible Cultural Heritage Digital Revitalization: A Conceptual Multi-Case Analysis 187
Mingjia Lin, Jianhong Sun, Chunqiong Wu and Tianhua Hui

The Happiness Backfire: Engagement-Oriented Brand Content Predicts Fewer Reposts and Shapes Comment Responses 196
Bowei Sun, Yanshu Wang and Tong Li

How Brand Happiness Spreads in User Co-Creation — Evidence from Large-Scale Social Media Data 203
Guo Tang, Yongtian Yu and Tong Li

Beyond Hashtag Quantity: How Content Types Trigger Customer Inspiration in Short-Video Marketing Differently 210
Lanxin Xie, Tong Li and Yongtian Yu

Research on the Impact of Digital-Intelligent Transformation on Corporate Performance of Manufacturing Enterprises 218
Lingling Hou and Jianhui Tan

Section 2. Track on "Applied-Education and Applied-Research in AI, Data Science, and Web6.0", the Second Edition

The Algorithmic Co-Thinker: A Generative AI Framework for Persona-Aware Computational Mathematics 231
Allan M. Canonigo

The Smart School Model: Integrating Intelligent Agents and IoT for Autonomous, Ethical and Reproducible Educational Management 240
Youssef Lahbabi, Abdellah Ait Oufkir, Lalla Touhfa Belgnaoui and Tarik Lafou

Towards a Click'N'Tick Global Culture: Inevitable Future in Digital Age 249
Ljupcho Efremov

An Automatic Evaluation Model for Health Science Popularization Short Videos Generated by Multimodal Large Language Models 258
Heshen Zhou, Yu Pei, Dongdong Zhang and Jing Wen

YouMii: A Daily Companion App for Mental Well-Being 264
Chun-Mun Look and Chien-Sing Lee

From Advice to Emergencies: Public Perceptions of AI Support in Social Media Service Requests 278
Mashael Almoqbel

Evaluating the Effectiveness of AI-Assisted Learning: An Empirical Study Based on Teaching Practice 290
Ruohan Li, Yongchang Wang, Lu Ma and Tong Xiao

Interactive Genetic Algorithm for Narrative-Driven VR Scene Layout Optimization 301
Yan Zhang, Wei Chen and Gang Yang

Lightweight Face Anti-Spoofing for Mobile Multi-Factor Attendance Systems 311
Ekkachan Rattanalerdnusorn and Sasakorn Pichetjamroen

Hybrid Financial Analysis Agent with RAG and Text Mining 321
Bingzhang Wang, Keyi Zhang, Tianhao Du, Xiaojie Chen, Jiyu Cao, Shaokang He and Adela S.M. Lau

ESG Chatbot: An Intelligent Platform for HKEX ESG Analysis 350
Sergey Volkov, Frederick Nathan Irmawan, Lei Lei, Kunheng Zuo, Pai Zhang, Duoyi Wu, Yunqi An and Adela S.M. Lau

A Retrieval-Based Multi-Source Medical Chatbot: Integrating Association Rule Mining and Bayesian Inference as a Health Information Support Tool 373
Peilin Luo, Anmaria Alice, Zongfu Chen, Jiayin Guo, Xinchu Zhou, Yutong Gu and Adela S.M. Lau

Humanizing Artificial Intelligence in Crisis: Digital Humans and Generative AI for Natural Disaster Management 400
Dimiter Velev and Plamena Zlateva

Integrating Lexicon-Based Approach and Machine Learning Method for Content Analysis: A Conceptual Framework 416
Sitong Li and Adela S.M. Lau

FinSight: An Evidence-First Financial Analysis Chatbot with Explainable Query Intelligence 424
Ruiqi Xie, Jiachang Zhang, Mingxi Shi, Shiyi Li, Mingyuan Gao, Zihao Yin and Adela S.M. Lau

Supermarket CRM Intelligent Analysis System: A Data Mining Approach for Customer Segmentation, Churn Prediction, and Personalized Recommendation 434
Yanting Guo, Zixin Wu, Yumeng Wang, Zhanhao Lei, Junyi Gao, Peizhen Lyu, Boyu Feng and Adela S.M. Lau

Thinking from Quantum Computing Theories to Problem and Applications by Using a Novel Ontology-Based Quantum Architecture Model 448
Yanyun Lu and Adela S.M. Lau

A Simulation-Based AI-Enabled Supply Chain Decision Framework for Cross-Border Business Expansion 468
Jiawei Qiu, Liyang Zeng, Zhuolun Wu, Chuyi Zou and Ben Fan

Research on Evaluation of Large Language Models for Healthcare 475
Peng Cheng, Hong Ruan, Xin Liao and Mingliang Tan

Section 3. Special Session on "New Ecology of Lifelong Education"

Mutual Information Analysis of Employment-Course Score Relationships for Curriculum Optimization in Higher Vocational Traditional Chinese Pharmacy 485
Ziquan Hu and Shuhao Hu

Construction of an Evaluation System for the Effectiveness of Innovative Spirit Education Among College Students Based on the Kirkpatrick Model: A Case Study of Higher Vocational Colleges in Guangdong Province 496
Yaojia Lin

Multi-Modal Digital Inheritance of Intangible Cultural Heritage: A Child-Friendly Micro-Course Framework and Empirical Study—Taking the "Xiangnan Little Heritage Bearers" Digital Curriculum System as an Example 508
Yan Peng, Han Wang and Di Wang

Research and Practice on the Construction of a Lightweight AI-Enabled Welding Experimental Teaching and Safety Intelligent Management System 515
Haifeng Yang, Gang Chen, Meirong Wang and Xiongyao Wang

Research on the Construction and Practice of AI Literacy Course 523
Xiaoli Liu, Wenqing Li, Jingjing Li, Zhen Zhang, Wei Li and Anjia Yang

AI Innovative Practical Teaching Model for Interdisciplinary Postgraduates Integrated with OBE Concept 531
Jiawei Qiu, Jianghui Liu, Yulin Zheng and Chuyi Zou

Predictive Validity of Online Learning Indicators for Collaborative Higher-Order Learning: Evidence from Learning Platform Data 538
Zixuan Huang, Chunqiong Wu and Jianhong Sun

Research on the Influencing Mechanism of Higher Vocational Students' Entrepreneurial Intention Based on Three-Dimensional Driving Mediation Model 547
Xiaoqi Huang and Shaozhong Zhang

Configurational Pathways in Digital Textbook Development: A TOE-Based Analysis of Business English Materials for Vocational Education 557
Jun Liu

Operational Indicator System for School-Enterprise Collaborative Virtual Teaching and Research Offices in Vocational Education: A CIPP Model-Based Study 575
Huanhuan Zhou

Section 4. Special Session on "The Potential of Big Data, Digitalization and AI in a Sustainable Economy"

AI-Driven Export Packaging Design: Challenges of Cultural Differences and Intelligent Response Strategies 586
Aihua Song

The Effects of Corporate Digital Transformation on Financial Information Governance: Evidence from Financial Restatements 593
Zhe Zhang and Ximeng Jia

Leap Forward or Proceed Step by Step? Regional Development Effects and Path Differentiation of China's New Quality Productivity 603
Yan-juan Hu

Configurational Pathways of Digital Transformation in New Energy Vehicle Manufacturing Enterprises 614
Jingrou Geng and Yibing Tang

Emission Reduction and Pricing Strategies in Low-Carbon Supply Chains Under Information Asymmetry and Risk Aversion 623
Yuxin Li and Yuxiang Yang

AI for Agriculture in China: Research Progress, Practical Dilemmas and Optimization Paths 632
Xiaoming Yan, Hailong Xia and Xicai Zhang

Cybersecurity Detection Strategies for the Digital-Intelligent Business Sand Simulation During the AI Generative Process 642
Guanghua Qiu, Zheng Wang, Guiqing Liu, Bier Liao and Jianghui Liu

The Dual Effects of Gamified Design on Gig Workers' Continuance Intention: A Multi-Path Examination Based on the S-O-R Framework 648
Yan Jin and Zanyu Chen

Modified Ratios for Assessing the Impact of Artificial Intelligence on the Economic Efficiency of Industrial Enterprises 662
Plamena Zlateva, Marin Galabov and Dimiter Velev

Digitalization and the International Competitiveness of Cultural Products: Evidence from China's Major Trading Partners 675
Yingyi Cao and Maoguang Pang

Breaking the "Last Mile" of Agricultural ESG: Evidence from 506 Smallholder Farmers 689
Jinfang Qian, Zixu Xia, Baijie Wang and Fang Guo

Section 5. Interdisciplinary Applications of Digitalization and Management Innovation

Research on the Influencing Factors of the Number of Collective Labor Disputes 700
Haichao Lu, Binbin Hu and Xiuyun Gong

Multi-Task Surface State Perception Enhanced by Self-Encoder-Decoder Features 718
Quan Gan, Qidan Zhu, Jing Li and Yang Tian

Artificial Intelligence in Discipline Group Development of Traditional Chinese Medicine Institutions: A Quantitative Systematic Review (2023-2025) 730
Ziquan Hu

Z-Number-Based Large-Scale Group Decision-Making Method Considering Conflict Behaviors 744
Yihan Yang, Tao Jiang, Zhujing Wang and Jingyi Sun

Carotid Plaque Prediction Based on Multi-Channel PPG Signals 751
Shuo Yuan, Guozhong Liu, Hengli Liu and Hao Meng

Wavefield Decomposition-Based Reverse Time Migration for Mirror Artifact Suppression in Horizontal Borehole Acoustic Imaging 758
Yonglong Yang, Haitao Yan, Zhao Le, Kaifeng Wu and Zhongsheng Yang

Leakage-Aware Road Damage Detection Based on GAM-YOLOv8 771
Jun Wang, Zuojun Bao, Wenjin Zeng, Chunting Qin and Jing Zhou

Physics-Guided Deep Learning for Rapid 1D TEM Inversion with Uncertainty Quantification 778
Zhao Le, Kaifeng Wu, Haitao Yan, Yonglong Yang, Hui Ye, Wentao Zhang, Shaobiao Liang and Depu Hu

Multiscale Progressive Pseudo-Label Self-Supervised Learning for End-to-End TEM Inversion 791
Zhao Le, Kaifeng Wu, Haitao Yan, Yonglong Yang, Hui Ye, Wentao Zhang, Shaobiao Liang and Depu Hu

Lightweight Chinese Medicinal Material Recognition via Coordinate Attention and Knowledge Distillation 803
Shuhui Min, Shaoyang Zheng, Tianwei Lei and Qinqin Wu

Research on an Intelligent Detection System for Municipal Manhole Covers Based on Edge-Cloud Collaboration 810
Haotian Luo, Zheng Sheng, Shengming Liang and Mingdi Li

Multi-Scale Dynamic Spatiotemporal Fusion for Traffic Flow Prediction 820
Zhiyong He and Xiongtao Zhang

Narcissistic Rivalry Effect on Employee Voice Behavior Under Abusive Supervision: A Moderated Mediation Model 829
Yunxia Su and Feng Li

ICT-Enabled Agricultural Systems for Food Security, Climate-Smart Agriculture, and Inclusive Rural Development 838
Rotimi-Williams Bello, Roseline Oluwaseun Ogundokun, Pius A. Owolawi, Etienne A. van Wyk, Chunling Tu and Odunayo Imoru

Load-Shedding-Resilient Precision Livestock Farming: Edge Deployment of a Real-Time Cattle Monitoring System for South African Dairy Farms 848
Thabo S. Makhorole, Chunling Tu, Pius A. Owolawi and Rotimi-Williams Bello

Study on FEM Text Cross-Platform Conversion and Intelligent Review Method 858
Jianyong Liu, Dong Fan, Huaiyu Tang, Yuteng Wen, Shaozhe Liu and Jiacheng Wei

Research on the Unsteady Aerodynamic Model of a Conical Probe Using Forced and Free Vibration Based on Parametric Dynamic Derivatives 871
Hao Han and Xu Liu

Multi-channel Audio Blind Source Separation and Enhancement System Based on Optimized FastICA 881
Xinyan Li, Yan Wang and Lanhua Zhang

Classification of Sandstone-Based Building Materials Based on YOLO Network 888
Yan Hong and Meng Lv

A Differential Interbeat Interval-Based Synchronization Method for Long-Term BCG and ECG Signals 895
Lijuan Yang, Haojun Huang, Han Zhang and Baoxian Yu

Aircraft Leasing Strategies in the Context of the Low-Altitude Economy via an Evolutionary Game and Simulation Approach 901
Wenchao Fang

Author Index 914

Section 1

Digitalization and Management Innovation

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261070

An Evaluation Study on the "Five-Dimensional Integrated" Digital Intelligence E-commerce Talent Cultivation Model Based on AHP-TOPSIS

Yan Wang[1]
Wuhan Huaxia Institute of Technology, China

Abstract. To address the dual challenges of talent shortages and the absence of robust evaluation frameworks in e-commerce training for rural revitalization, this study develops a "Five-Dimensional Integrated" model for cultivating digital-intelligent e-commerce talent. Integrating the Analytic Hierarchy Process (AHP) and the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS), we establish a comprehensive evaluation system grounded in a "theoretical preset, empirical measurement, cross-validation" logic. Expert consultations informed the weighting of evaluation indicators, enabling a quantitative assessment and statistical comparison between an experimental group (exposed to the new model) and a control group (following the traditional approach). The results indicate that the average relative proximity score of the experimental group (0.684) significantly exceeds that of the control group (0.326). Furthermore, a significant positive correlation exists between the AHP-derived indicator weights and the relative performance advantage of the experimental group ($\rho = 0.82, p < .05$), demonstrating that the cultivation model most effectively enhances the specific competencies prioritized by expert consensus. Thus, this study not only validates the efficacy of the proposed training model but also delivers a quantifiable and diagnostic toolkit for evaluating and improving talent cultivation outcomes.

Keywords. Digital and Intelligent E-commerce Talent; Rural Revitalization; AHP-TOPSIS; Training Model; Comprehensive Evaluation

1 Problem Statement and Research Design

1.1 Research Background

Despite being a pivotal engine for industrial upgrading and economic growth within China's national rural revitalization strategy, the sustainable development of rural e-commerce is constrained by a critical and persistent bottleneck: a structural mismatch in human capital [1, 2]. This challenge extends beyond a mere quantitative deficit. Industry forecasts consistently project a widening talent gap in the coming years [3, 4]. More fundamentally, it is a qualitative crisis of competence. Graduates from conventional training programs frequently lack the integrated skill set required by the modern digital

[1] Corresponding Author, Yan Wang, Wuhan Huaxia Institute of Technology, China; Email: mandy9905303@126.com

marketplace, exhibiting underdeveloped technical expertise, superficial understanding of agricultural value chains, and deficient practical operational experience [5, 6]. This misalignment between talent supply and industry demand threatens to undermine the long-term efficacy of substantial investments in rural digital infrastructure and policy frameworks.

1.2 Literature Review

To establish a rigorous theoretical foundation and construct a robust initial pool of evaluation indicators, a systematic review of the literature was undertaken. A targeted search was conducted using major academic databases, including the China National Knowledge Infrastructure (CNKI) and the Web of Science Core Collection. Key search terms encompassed "rural revitalization," "e-commerce talent," "cultivation model," "competency evaluation," "AHP," and "TOPSIS." The review focused on peer-reviewed journal articles published between 2018 and 2024, prioritizing studies that explicitly addressed the development or assessment of talent within the rural e-commerce ecosystem. The extant research can be categorized into three interrelated streams. The macro-level stream investigates institutional environments and systemic barriers. For example, studies have analyzed structural impediments to e-commerce's role in rural development [5, 6] and imbalances in the distribution of critical support talent, such as in logistics [7]. The meso-level stream focuses on pedagogical and institutional innovation in training models. Contributions here include frameworks for multi-stakeholder cultivation systems grounded in institutional economics [8] and dynamic pedagogical paradigms that integrate technical skills with broader cultural and innovative competencies [9]. Finally, the micro-level stream deconstructs the specific competency architecture required for professional practice, defining core skill sets and knowledge domains for rural e-commerce practitioners [10].

However, existing research exhibits a significant disconnect in the transition from "theoretical construction" to "practical validation" and "continuous optimization." Most studies remain confined to qualitative analysis (Han Yujie, 2023) or case-based experience summaries (Wang Wei, 2022). While these approaches have accumulated valuable practical insights, they struggle to establish a set of quantifiable, scalable, and comparable universal evaluation standards [11,12]. More critically, despite consensus on the necessity of quantitative evaluation, existing assessment frameworks often lack a solid methodological foundation. In the core process of determining indicator weights—which dictates the scientific rigor of evaluation—most studies rely on subjective experience-based weighting. Rarely do they employ mature decision models such as the Analytic Hierarchy Process (AHP), resulting in insufficient reliability and validity of assessment outcomes [13,14,15].

1.3 Research Questions and Significance

Grounded in the preceding analysis, this study seeks to address the core research question: How can the effectiveness of a digital-intelligent e-commerce talent cultivation model be scientifically evaluated and validated? This overarching question is operationalized through three specific sub-questions:

(1) Indicator System Development: How can a systematic, scientifically grounded, and rural revitalization-aligned comprehensive evaluation indicator system be constructed to assess the outcomes of digital-intelligent e-commerce talent cultivation?

(2) Methodological Integration: How can the Analytic Hierarchy Process (AHP) and the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) be integrated to develop an empirical evaluation model that objectively derives indicator weights and precisely quantifies the relative performance of different training outcomes?

(3) Model Efficacy and Validation: Employing this integrated model, does the "Five-Dimensional Integrated" digital-intelligent cultivation model demonstrate statistically significant superiority over traditional approaches in enhancing students' composite capabilities? Furthermore, is there congruence between the model's theoretical priorities (as reflected in AHP weights) and its empirical performance outcomes?

In theoretical terms, this paper integrates the AHP and TOPSIS models to organically combine expert qualitative consensus (theoretical assumptions) with objective performance data (empirical measurements), establishing a complete methodological closed-loop of "assumption-measurement-verification" [14,15]. Simultaneously, the study systematically incorporates cutting-edge digital intelligence capability dimensions (such as data literacy and digital thinking) into traditional e-commerce talent evaluation systems, enriching the theoretical framework of the field.

In practical terms, the developed "five-dimensional integrated" training model and AHP-TOPSIS evaluation tool provide universities and enterprises with a clear reform roadmap and precise diagnostic instruments. Evaluation outcomes not only facilitate comparative analysis of different models but also pinpoint specific strengths and weaknesses in the training process, offering direct data support for curriculum adjustments and resource optimization.

2. Development of the "Five-Dimensional Integrated" Training Model and Evaluation System

2.1 The Conceptual Framework of the "Five-Dimensional Integration" Model

This study introduces a "Five-Dimension Integrated" cultivation model that establishes a comprehensive framework for talent development. This organic system is built upon a foundation of digital thinking, extended through data literacy enhancement, structured around digital-intelligence curricula, supported by digital-intelligence platforms, and ultimately directed toward developing digital-intelligence talent.

As illustrated in Figure 1, the model features three interconnected tiers: the Value Shaping tier cultivates data-driven cognitive frameworks; the Competency Development tier facilitates the translation of theoretical knowledge into practical capabilities through synergistic integration of data literacy, digital-intelligence curricula, and platform support; and the Goal Realization tier produces versatile professionals equipped to manage complex rural e-commerce operations.

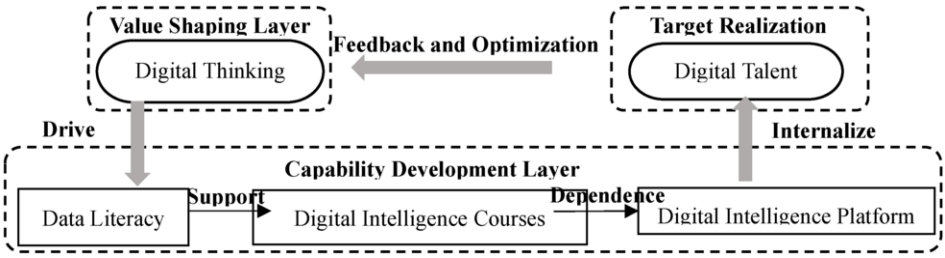


Figure 1: "Five-Dimensional Integrated" Digital Intelligence E-commerce Talent Development Model

The model's distinctive innovation resides in its embedded dynamic adaptation mechanism. This self-adjusting system leverages industry feedback from professionals to iteratively refine both the curricular content of digital-intelligence courses and the resource configuration of supporting platforms. Concurrently, it progressively enriches the conceptual framework of digital thinking, thereby establishing a self-improving closed-loop ecosystem. This integrated mechanism ensures continuous alignment between the talent cultivation process and evolving industry requirements, enabling responsive adaptation to the dynamic landscape of rural e-commerce development.

2.2 Design of the Comprehensive Evaluation Indicator System

The "Five-Dimensional Integrated" evaluation framework is theoretically grounded in the systematic integration of pedagogical principles, industry requirements, and methodological rigor. The framework comprises five core dimensions—digital thinking, data literacy, digital-intelligence curricula, digital-intelligence platforms, and digital-intelligence talent—which are operationalized through 13 quantifiable secondary indicators. These measurable metrics include: data-driven decision-making frequency, data analysis report quality, 1+X certification attainment rates, practical project transaction volume, and agricultural sector employment/entrepreneurship rates. Together, these components form a comprehensive and systematically quantifiable assessment repository for evaluating talent development effectiveness.

2.3 Establishment of the AHP-TOPSIS Evaluation Model

This paper constructs an integrated evaluation model based on AHP (Analytic Hierarchy Process) and TOPSIS (Technique for Order Preference by Similarity to Ideal Solution). The core logic of this model is to establish a progressive analytical framework of "theoretical assumptions—empirical measurements—cross-validation," as illustrated in Figure 2.

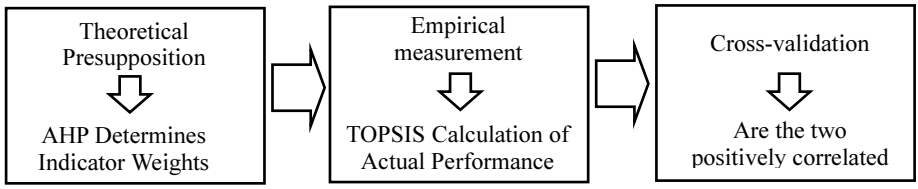


Figure 2: Framework of the AHP-TOPSIS Evaluation Model

2.3.1 AHP Weight Determination

This study employs the Analytic Hierarchy Process (AHP) to determine evaluation indicator weights through a structured hierarchical framework consisting of objective, criterion, and alternative layers. Fifteen domain experts from academic e-commerce departments, rural revitalization research institutions, and e-commerce enterprises participated in pairwise comparisons using Saaty's 1-9 scaling method. The weight calculation process involved eigenvector methodology and consistency validation ($CR < 0.1$), resulting in a robust weighting system that demonstrates satisfactory consistency.

2.3.2 Ranking using TOPSIS

In this phase, the TOPSIS method was employed to quantitatively measure and rank the actual performance of the experimental and control groups based on the weighting scale provided by AHP. The operational procedure is as follows:

- Collect raw data for both student groups across all secondary indicators to construct the initial decision matrix.
- Perform vector normalization on the matrix to eliminate dimensionality, then multiply by the comprehensive weights determined by AHP to form a weighted normalized decision matrix. This step ensures measurements reflect both objective performance and theoretical significance.
- Identify the positive ideal solutions (optimal value set) and negative ideal solutions (worst value set) for each indicator in the weighted matrix.
- Calculate the Euclidean distance between each evaluation object and its positive and negative ideal solutions (denoted as D^+ and D^-, respectively).
- Compute the final score for each object using the relative proximity formula $C = D^-/ (D^+ + D^-)$. A higher C value (closer to 1) indicates that the object's overall performance is closer to the ideal state. By comparing the distribution and ranking of C values across groups, an objective quantitative assessment of training effectiveness can be achieved.

2.3.3 Cross-validation

If the experimental group's key indicators with higher AHP weights also demonstrate significantly superior actual performance measured by TOPSIS compared to the control group, this indicates a strong positive correlation between theoretical importance (AHP weight) and actual superiority (TOPSIS proximity). Such findings would provide compelling evidence that the "five-dimensional integrated" cultivation model not only aligns with expert expectations but also yields remarkable implementation outcomes.

3. Empirical Analysis and Results

3.1 Data Sources and Processing

This study employed a quasi-experimental design, targeting two parallel classes of the 2022 cohort majoring in E-commerce at Wuhan Huaxia University of Science and

Technology. With individual students as the basic unit of analysis, a two-year longitudinal study was conducted. To ensure internal validity, rigorous baseline homogeneity tests were performed on the experimental group (Class A, 26 students) and the control group (Class B, 24 students) prior to the experiment. We systematically compared differences between the two groups across key background and academic indicators, including total entrance examination scores, gender composition, place of origin (urban/rural), and final grades in the core course "Introduction to E-Commerce" during the first semester. The results of the baseline homogeneity tests are as follows:

Total entrance exam scores: An independent samples t-test revealed no significant difference between the two groups ($t(48) = 0.42$, $p = .676$).

Gender Composition: A chi-square test revealed no significant difference in gender distribution between the two groups ($\chi^2(1) = 0.12$, $p = .735$).

Origin Distribution: A chi-square test showed no significant difference in the proportion of urban/rural origin between the two groups ($\chi^2(1) = 0.08$, $p = .777$).

Prerequisite Course Grades: An independent samples t-test revealed no significant difference in grades for the Introduction to E-Commerce course between the two groups ($t(48) = 0.85$, $p = .401$).

The above results indicate that the experimental and control groups were comparable on all baseline characteristics measured ($p > .05$), satisfying the homogeneity requirement for quasi-experimental designs.

Data were collected through multiple validated sources to ensure comprehensive coverage: academic performance records were extracted from the institutional student information system, certification attainment was verified through the National Vocational Skills Qualification Platform, competition outcomes were obtained from the University Competition Management Database, and practical project metrics were derived from school-enterprise collaboration platform interfaces. Employment tracking data were sourced from the official Graduate Destination Survey System. To enhance data reliability and validity, this study implemented methodological triangulation, cross-validating information across these diverse data streams. For data standardization, min-max normalization was applied to transform all raw indicators to a standardized [0,1] scale.

3.2 AHP Weight Analysis

The weight distribution derived through the Analytic Hierarchy Process for the "Five-Dimension Integrated" model (presented in Table 1) reveals distinct priorities in digital-intelligent e-commerce talent cultivation. All computations in this phase were performed in a Python 3.9 environment, utilizing the numpy library for feature vector calculations and invoking the scipy library to assist in verifying the consistency ratio (CR). The calculated consistency ratios (CR) for all expert judgment matrices were found to be less than 0.1 (specifically ranging from 0.02 to 0.08), meeting the consistency requirements. This indicates that the weighting results are valid and reliable.

Table 1 Weight Distribution of Quantitative Evaluation Indicators for the "Five-Dimensional Integrated" Model

Primary Indicator (Guideline Level)	Weight	Secondary Indicators (Scheme Level)	Relative Weight	Composite Weighting	Sorting
Digital ThinkingA_1	0.152	Data-Driven Decision RateA_{11}	0.400	0.061	10
		Frequency of operational strategy iterationsA_{12}	0.350	0.053	11
		Number of Collaborative Solutions for Industrial ChainsA_{13}	0.250	0.038	13
Data LiteracyA_2	0.198	Data Analysis Report QualityA_{21}	0.600	0.119	2
		Proficiency in Data ToolsA_{23}	0.400	0.079	6
Digital Intelligence CoursesA_3	0.165	Advanced Course GradesA_{31}	0.500	0.083	4
		"1+X"Certificate Acquisition Rate A_{32}	0.500	0.083	4
Digital Intelligence PlatformA_4	0.241	Simulation Platform Completion RateA_{41}	0.200	0.048	12
		Actual Project Transaction VolumeA_{42}	0.500	0.121	1
		Platform Task Response TimeA_{43}	0.300	0.072	9
Digital TalentA_5	0.244	Professional Competition Award LevelsA_{51}	0.300	0.073	7
		Employment/Entrepreneurship Rate in Agriculture-Related FieldsA_{52}	0.400	0.098	3
		Employer SatisfactionA_{53}	0.300	0.073	7

The weight analysis reveals a distinct prioritization among evaluation criteria, with "Digital-Intelligent Talent" (0.244) and "Digital-Intelligent Platforms" (0.241) demonstrating significantly higher weights than other dimensions. Together constituting 48.5% of the total weighting, this distribution highlights the framework's dual emphasis on both talent development outcomes and practical implementation infrastructure.

To empirically examine the rationality of AHP weight allocation and its correlation with training effectiveness, this study further calculated the Spearman rank correlation coefficient between the weight of primary indicators and the improvement margin of the experimental group relative to the control group on that indicator (experimental group mean/control group mean). Results revealed a significant positive correlation ($\rho = 0.82$, $p = .045$), indicating that dimensions assigned higher weights by experts corresponded precisely to areas where the experimental group achieved greater relative advantages. This empirically supports the validity of the weighting design.

At the secondary indicator level, "Practical Project Transaction Volume" (0.121) emerges with the highest individual weight, reinforcing the model's commitment to outcome-based education and tangible skill development. This weighting pattern reflects three strategic priorities: the dominance of practical application capabilities (0.193 combined weight for hands-on skill indicators), the foundational importance of data literacy (0.198 combined weight), and clear alignment with rural revitalization objectives through "Agricultural Sector Employment/Entrepreneurship Rate" (0.098).

3.3 TOPSIS Comprehensive Evaluation Results

This paper employs Python 3.9 to construct a TOPSIS evaluation model, utilizing the pandas and numpy libraries to generate the weighted normalized matrix, determine positive and negative ideal solutions, and compute Euclidean distances and proximity values (C-values). The model performs comprehensive calculations on all 50 student

subjects. To clearly illustrate computational logic and individual variation, Table 2 presents detailed data for 10 representative samples—5 randomly selected from each group.

Table 2 TOPSIS Comprehensive Evaluation Results (Representative Sample)

Sample Number	Group	Distance from Positive Ideal Solution (D^+)	Distance from Negative Ideal Solution (D^-)	Relative Proximity (C-value)	Overall Ranking
S1	Experimental group	0.127	0.473	0.788	1
S2	Experimental group	0.152	0.458	0.751	2
S3	Experimental group	0.185	0.425	0.697	3
S4	Experimental group	0.241	0.389	0.617	4
S5	Experimental group	0.283	0.367	0.565	5
C1	Control group	0.398	0.282	0.415	6
C2	Control group	0.425	0.255	0.375	7
C3	Control group	0.463	0.227	0.329	8
C4	Control group	0.502	0.198	0.283	9
C5	Control group	0.547	0.163	0.230	10

The comprehensive analysis results based on the entire sample are as follows: The experimental group had an average proximity of 0.684 (SD = 0.085), while the control group had an average proximity of 0.326 (SD = 0.073). To test the statistical significance of the difference in overall performance between the two groups, a Mann-Whitney U test was conducted on the proximity data of all 50 students. The results indicated that the experimental group's overall performance was significantly superior to that of the control group ($U = 89$, $Z = -3.92$, $p < .001$), with a large effect size ($r = .78$).

Further examination of core process indicators in TOPSIS calculations:

Distance to Positive Ideal Solution (D^+): The experimental group's mean D^+ (0.198) was significantly smaller than the control group's (0.467). Independent samples t-test results: $t(48) = 15.32$, $p < .001$.

Distance from Negative Ideal Solution (D^-): The experimental group's mean D^- (0.422) was significantly greater than the control group's (0.225). Independent samples t-test results: $t(48) = 10.47$, $p < .001$.

These results indicate that students in the experimental group were overall closer to optimal performance while being farther from suboptimal performance.

The analysis reveals three key findings regarding the model's efficacy. First, the experimental group achieved a mean proximity score of 0.684—2.1 times higher than the control group's 0.326—demonstrating substantial improvement in overall talent competency. Second, the experimental group exhibited greater internal consistency (C-values ranging 0.565–0.788) compared to the wider dispersion observed in the control group (0.230–0.415), indicating more stable and predictable training outcomes. Most notably, the lowest-performing experimental case (S5, 0.565) exceeded the highest-performing control case (C1, 0.415), providing statistical evidence of the model's comprehensive superiority across the performance spectrum.

3.4 Cross-validation of theoretical assumptions and empirical outcome

This paper further examines the relationship between the theoretical weights determined by AHP and the empirical results of TOPSIS measurements.

First, validation was conducted at the dimension level. The AHP weights for the five primary indicators (A1-A5) were calculated and subjected to Spearman's rank

correlation analysis with the mean standardized TOPSIS scores of all secondary indicators under that dimension in the experimental group (representing the actual cultivation effectiveness of that dimension). The results revealed a significant positive correlation ($\rho = 0.82$, $p = .045$). This indicates that dimensions assigned higher weights by experts—such as "Digital Intelligence Platform" and "Digital Intelligence Talent"—correspond precisely to areas where experimental group students demonstrated superior performance. This confirms, at the dimensional level, that the design priorities of the cultivation model align strongly with its actual outcomes.

Second, validation was conducted at the key indicator level. Focusing on the top five key secondary indicators ranked by composite weight (see Table 1), the TOPSIS standardized scores for these indicators were compared between the experimental and control groups. As shown in Table 3, the experimental group scored significantly higher than the control group on all five key indicators ($p < .01$ via independent samples t-test). Notably, on the highest-weighted indicator "Practical Project Transaction Volume (A42)," the experimental group's score was 2.2 times that of the control group. This directly demonstrates that the model most effectively enhances the practical and output capabilities deemed most critical by experts.

Table 3: Comparison of Key Secondary Indicators Between Experimental and Control Groups

Sorting	Secondary Indicators	Comprehensive Weighting	Experimental Group Average Score	Control Group Average Score	Intergroup Difference Detection (Independent Samples t-Test)	t-test p-value
1	Actual Project Transaction Volume (A_{42})	0.121	0.85	0.39	t(48)=8.91	<0.001
2	Data Analysis Report Quality (A_{21})	0.119	0.78	0.42	t(48)=6.45	.002
3	Employment/Entrepreneurship Rate in Agriculture-Related Fields (A_{52})	0.098	0.80	0.45	t(48)=5.89	.003
4	Advanced Course Grades (A_{32})	0.083	0.82	0.76	t(48)=3.12	.003
5	"1+X" Certificate Acquisition Rate (A_{31})	0.083	0.88	0.65	t(48)=3.12	.003

Note: Scores represent standardized values [0,1] prior to TOPSIS processing.

Finally, we conducted an overall explanatory power test for the model. Using the AHP composite weights as the independent variable and the improvement in TOPSIS scores for each student in the experimental group on corresponding indicators (experimental group score minus control group average score) as the dependent variable, we performed weighted least squares regression analysis to examine the extent to which the weights explain the improvement in outcomes. The model fit results ($R^2 = 0.71$, $F = 25.34$, $p < .001$) indicate that the AHP weighting system effectively explains the majority (71%) of the variance in individual student capability improvement. This provides strong statistical support for the core logic that "theoretical assumptions drive empirical outcomes."

4. Discussion and Recommendations

4.1 Discussion of Results

This paper systematically demonstrates the effectiveness of the "five-dimensional integrated" digital and intelligent e-commerce talent cultivation model by constructing a comprehensive analytical framework encompassing theoretical assumptions, empirical measurement, and cross-validation. Evidence supporting this model was obtained from three levels: the internal logic of the evaluation model, statistical inference, and theoretical-empirical correlation.

First, the model demonstrates significant effectiveness. TOPSIS analysis reveals that the experimental group's overall performance significantly outperforms the control group (average proximity ratio of 2.1:1, U-test $p < .001$, large effect size). The experimental group not only approached the "optimal solution" more closely (smaller D^+) but also distanced itself further from the "worst solution" (larger D^-). This confirms the comprehensive superiority of its cultivation effects within the geometric space of multi-criteria decision-making, rather than merely leading in a single metric.

Second, the statistical inference results are robust. Beyond the significant difference in proximity (C-value), independent samples t-tests for key secondary indicators (as shown in Table 3) further confirm that the experimental group established statistically significant advantages across expert-identified core competency items—such as actual project transaction volume and data analysis report quality. This indicates that the model's effectiveness is systemic rather than attributable to individual students' heterogeneous performance.

Finally, theoretical assumptions are highly correlated with empirical outcomes. The most methodologically significant finding of this study is that Spearman's correlation analysis confirmed a significant positive correlation ($\rho = .82$, $p = .045$) between the theoretical weights determined by AHP and the relative advantage levels of the experimental group. This finding directly links expert theoretical consensus with the measurement of a training model's "actual effectiveness." It provides compelling evidence that the "Five-Dimensional Integration" model successfully directs training resources and instructional focus precisely toward high-value competency domains jointly identified by industry and academia. This effectively resolves the fundamental disconnect between 'teaching' and "industry needs" inherent in traditional training approaches [2,8].

4.2 Practical Insights and Recommendations

4.2.1 Establish a Dual-Core Driving Mechanism of "Platform-Talent"

Empirical findings indicate that the "digital intelligence platform" and "digital intelligence talent" are the two most critical dimensions determining training effectiveness (combined weight: 48.5%). Universities should establish a dual-core investment and development strategy: Regarding platform construction, efforts must transcend simple software procurement or laboratory establishment, focusing instead on building an integrated immersive teaching environment that combines simulation, practical application, and iteration. This ensures students can hone their comprehensive

operational capabilities—represented by "actual project transaction volume"—in near-real environments. Regarding talent output, universities should establish a comprehensive tracking and service system guided by "employment/entrepreneurship rates in agriculture-related fields" and "employer satisfaction." By deepening order-based training with rural e-commerce enterprises and cooperatives, and co-building industry colleges, the talent cultivation process should be seamlessly aligned with the job demands of rural revitalization [6,7].

4.2.2 Deepening Holistic Development of Data Literacy

Data literacy is the core manifestation of digital intelligence capabilities (weighting 0.198). Its cultivation should not rely solely on one or two standalone courses but should be integrated throughout the entire professional education process [10]. In terms of curriculum systems, promote the "digital intelligence transformation" of all core professional courses, deeply integrating data collection, analysis, visualization, and decision-making thinking and tools into traditional teaching content such as marketing, supply chain management, and customer service [9]. Regarding the certification system, vigorously advance the deep integration of the "1+X" certificate system with the credit bank, encouraging students to obtain professional skill level certificates such as Data Analyst and E-commerce Data Analyst, and convert these into innovation practice credits[11].

4.2.3 Establishing a Continuous Improvement System Based on Quantitative Evidence

(1) Establish an annual talent development quality diagnostic reporting system. Regularly apply this model to evaluate current students and graduates, generating quantitative diagnostic reports that include comprehensive alignment scores, dimensional strengths and weaknesses, and comparisons with previous cohorts [5].

(2) Form a closed-loop management system of "evaluation-feedback-optimization." Leverage diagnostic findings to revise talent development plans, adjust course content, optimize practical training project allocation, and reallocate teaching resources [9].

(3) Integrate diverse, long-term feedback channels. Combine model evaluations with practical assessments from industry mentors, graduate career development tracking surveys, and industrial technology trend reports to create a dynamic, multifaceted, and sustainable ecosystem for monitoring and improving talent development quality [12].

5. Conclusions

5.1 Research Summary

This study systematically investigates innovative approaches for cultivating e-commerce talent within the rural revitalization context, proposing a "Five-Dimension Integrated" digital-intelligent e-commerce talent development model and its corresponding AHP-TOPSIS comprehensive evaluation framework. Empirical results indicate that the model significantly enhances students' comprehensive competencies, with the experimental

group achieving an average proximity score 2.1 times higher than the control group, particularly excelling in indicators reflecting digital-intelligent capabilities.

The research validates that the "Five-Dimension Integrated" model—grounded in digital thinking, expanded through data literacy, centered on digital-intelligent curricula, supported by digital-intelligent platforms, and directed toward digital-intelligent talent development—effectively addresses the theory-practice disconnection and the competency-demand mismatch prevalent in conventional training methods. Moreover, the AHP-TOPSIS evaluation framework not only offers a scientifically robust methodology for assessing talent cultivation quality but also establishes a quantifiable, comparable, and diagnostically capable monitoring mechanism. This system provides a reliable foundation for continuous model refinement and sustainable quality improvement.

5.2 Research Limitations and Future Prospects

While this study has achieved its primary objectives in model development and validation, several limitations merit consideration. The research sample, drawn exclusively from institutions within a single geographical region, constrains both sample diversity and generalizability. Furthermore, although the evaluation framework encompasses five key dimensions, it inadequately captures certain qualities essential in rural contexts—such as humanistic literacy and sustainable development competencies. The reliance on cross-sectional data also limits insights into the dynamic processes of talent development.

To address these constraints, three promising research directions emerge. First, expanding sample diversity through cross-regional and multi-institutional comparisons would enhance the model's robustness and applicability. Second, integrating qualitative dimensions—including rural cultural identity, digital ethics, and social responsibility—would yield a more holistic assessment framework. Third, longitudinal tracking of students' developmental trajectories would illuminate the dynamics of digital-intelligent competency formation. Additionally, applying this evaluation model to adjacent domains—such as digital cultural tourism and smart agriculture—could reveal its broader utility within new liberal arts education.

Acknowledgment

This paper is a phased outcome of the 2024 Hubei Provincial Teaching Reform Research Project "Research on the 'Five-Dimension Integrated' Digital-Intelligent Innovation Talent Cultivation Model for E-commerce Majors Driven by Rural Revitalization" (2024551) and the 2023 Hubei Provincial Outstanding Middle-aged and Young Scientific and Technological Innovation Team Project (T2023049).

References

[1] Liu, X. Cui. Research on E-commerce Empowering Rural Economic Development under the Rural Revitalization Strategy [J]. Shanghai Business, 2024, (05): 60-62.

[2] Cheng, C. Zong, Wang, Y. N., Meng, D. Y. Research on E-commerce Empowering Rural Economic Development under the Rural Revitalization Strategy [J]. China Market, 2023, (30): 69-72.

[3] Dong Yun. Research on Establishing an E-commerce Poverty Alleviation Effectiveness System Under the Rural Revitalization Strategy [J]. Modern Market, 2023, (02): 42-44.
[4] Ren Yedi, Zhang Xinru. Research on Issues and Countermeasures for E-commerce Assistance to Agriculture Under the Rural Revitalization Strategy [J]. Commercial Exhibition Economy, 2022, (14): 65-67.
[5] Zhan Luyao, Han Jianmin, Su Xudong. Talent Demand Challenges and Cultivation Pathways for Rural E-commerce Logistics Development under the Rural Revitalization Strategy [J]. Logistics Engineering and Management, 2025, 47(06): 64-67.
[6] Chen Fengjuan. A Brief Analysis of Problems and Countermeasures in E-commerce Assistance for Agriculture under the Rural Revitalization Strategy [J]. Modern Business Research, 2024, (15): 34-36.
[7] Hua Xia, Tang Qingchao. Problems and Countermeasures in E-commerce Assistance for Agriculture under the Rural Revitalization Strategy [J]. Journal of Shandong Agricultural Engineering University, 2022, 39(12): 44-48.
[8] Yu Bo. Innovation and Pathways for Rural E-commerce Talent Development Models under the Rural Revitalization Strategy [J]. Agriculture and Technology, 2025, 45(10): 126-129.
[9] Mo Haiyan, Mo Limei, Li Xiang. Practical Research on Rural E-commerce Talent Development under the Rural Revitalization Strategy [J]. Guangxi Education, 2022, (23): 122-126.
[10] Hou Xufang. Research on Building a Competency Development System for Rural E-commerce Professionals Under the Rural Revitalization Strategy [J]. Shanxi Agricultural Economics, 2022, (12): 159-161.
[11] Han Yujie. Research on the Current Status and Cultivation Strategies of Rural E-commerce Talent Under the Rural Revitalization Strategy [J]. Modern Agriculture, 2023, (09): 78-80.
[12] Wang Wei. Research on Training Strategies for Rural E-commerce Talent under the Rural Revitalization Strategy [J]. Public Relations World, 2022, (21): 44-45.
[13] Huang Dongmei, Lu Huanhuan, Liu Yin, et al. Safety Evaluation of University Teaching Laboratories Based on the AHP-TOPSIS Model [J]. Science, Education & Culture, 2025, (18): 30-34.
[14] Chen Junying, Xing Yuhu, Li Jing, et al. An In-Depth Exploration of the QEABM Higher Education Evaluation System Based on Talent Cultivation [J]. Journal of Xihua Normal University (Natural Science Edition), 2020, 41(03): 331-340.
[15] Li Ang, Sun Wei. Evaluation of Industry-Academia-Research Collaboration Outcomes in Shandong Higher Education Institutions Using the AHP-TOPSIS Model [J]. Heilongjiang Science, 2023, 14(11): 74-76+82.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)

© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261071

The Adoption and Impact of Smart Wearable Devices: A Literature Review Based on the "Antecedent-Mechanism-Consequence" Framework

Yawei Wang[a, b] and Qinmei Wu[a,1,2]
[a] *School of Management, Wuhan Textile University, Wuhan 430200, Hubei Province, China*
[b] *Hubei Key Research Base of Humanities and Social Sciences in General Colleges and Universities-Enterprise Decision Support Research Center, Wuhan 430073, Hubei Province, China*
ORCiD ID: Wu Qinmei https://orcid.org/0009-0000-8043-5669

Abstract. Smart wearable devices have rapidly evolved from niche self-tracking tools to integral components of digital health ecosystems. Although prior research has examined adoption drivers and usage outcomes, the literature remains fragmented, with limited integration of antecedents, mediating mechanisms, and multilevel consequences. To address this gap, this study conducts a systematic review of 5,501 Web of Science publications (2010–2025) following PRISMA guidelines and employs CiteSpace co-citation analysis to map the intellectual structure of the field. Building on the Antecedent–Mechanism–Consequence (AMC) framework, the review synthesizes key motivational, psychological, functional, and contextual drivers of wearable use and identifies the mediating processes through which these factors translate into individual, organizational, and societal outcomes. The findings reveal a patterned ambivalence in wearable effects, characterized by self-regulation–dependence tension at the individual level, personalization–privacy trade-offs at the organizational level, and efficiency–surveillance dilemmas at the societal level. This study advances existing literature by providing an integrated process view of smart wearable use and proposing a theory-driven research agenda to guide future work on sustainable and human-centered wearable ecosystems.

Keywords. smart wearable devices; literature review; integrated framework; research prospects

[1] Correspond Author, Qinmei Wu, Wuhan Textile University, China; Email: 17762978051@163.com
[2] Project supported by: 1) Hubei Provincial Department of Education Philosophy and Social Sciences Research Program(Grant No.25Q092) ; 2) Research Center of Enterprise Decision Support, Key Research Institute of Humanities and Social Sciences in Universities of Hubei Province(Grant No.DSS20240708); 3) Wuhan Textile University Special Fund (Grant No.2024387) ; 4) Wuhan Textile University Fund (Grant No.20230612).

1. Introduction

Smart wearable devices have become deeply embedded in digital health and everyday self-tracking practices, enabling continuous monitoring and personalized behavioral support. Existing studies have examined various adoption drivers and usage outcomes; however, the literature remains fragmented, with limited understanding of how antecedent conditions translate into multilevel consequences through underlying mechanisms.

Most prior work focuses on isolated predictors or specific health outcomes, providing insufficient process-level explanation of when wearable technologies generate sustained value versus unintended negative effects such as dependence, privacy concerns, and motivational fatigue. Moreover, few reviews offer a theory-driven synthesis that systematically connects drivers, mediating pathways, and consequences within a unified analytical structure.

To address this gap, this study conducts a PRISMA-guided systematic review of smart wearable research and develops an Antecedent–Mechanism–Consequence (AMC) framework to organize the field. Based on bibliometric mapping and thematic synthesis, the study identifies key antecedents, clarifies mediating processes, and examines multilevel outcomes of wearable use.

This review contributes by providing an integrated process view of wearable adoption and proposing a research agenda for more sustainable and human-centered wearable ecosystems.

2. Literature Review Foundation

2.1 Smart Wearable Devices: Definition and Scope

Smart wearable devices are body-worn electronic systems that enable continuous sensing, data processing, and real-time feedback in everyday contexts. Prior research highlights their always-on accessibility, embedded interaction capability, and role in personal health and behavior monitoring. Compared with medical-grade monitoring equipment and home health systems, smart wearables are distinguished by portability, user-centered design, and voluntary everyday use.

This study defines smart wearables as intelligent wearable technologies that collect and analyze physiological or behavioral data while providing timely feedback to users in daily life. This definition clarifies the conceptual boundary and provides the basis for subsequent analysis.

2.2 Systematic Review Methodology

This study followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines to ensure transparency and reproducibility. The literature search was conducted in the Web of Science Core Collection using the keyword "smart wearable device". The search covered publications from 2010 to 2025, retaining only peer-reviewed English articles. After duplicate removal and multi-stage screening based on inclusion and exclusion criteria, 5,501 articles were included for

bibliometric and thematic analysis using CiteSpace. The selection process is illustrated in Figure 1

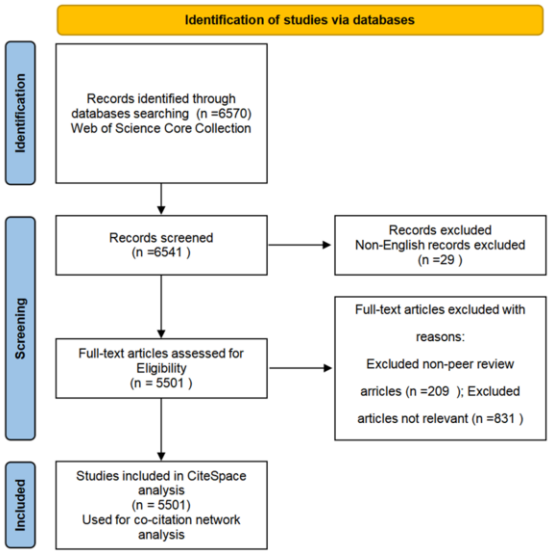


Figure 1 PRISMA flowchart of the article selection process

2.3 Development History and Research Evolution of Smart Wearable Devices

To reveal the developmental trajectory of smart wearables, a co-citation network analysis was conducted using CiteSpace on 5,501 Web of Science records. As shown in Figure 2, the knowledge structure forms several major clusters, including conductive hydrogel, ambient intelligence, and smart wearable applications, reflecting shifts in research focus over time.

The field can be divided into three stages: technological emergence (1970s–2010), human–computer interaction transformation (2011–2016), and cross-domain integration (2017–present). Early research addressed core technical feasibility, while the second stage emphasized interactive intelligence and self-tracking capabilities. Recent studies highlight interdisciplinary integration and multi-scenario applications.

The bibliometric evidence indicates a transition from foundational technology development to intelligent interaction integration.



Figure 2 Use CiteSpace to visualize the co-citation reference network (2000-2025)
Note: Cluster visualization of co-citation networks. Node size indicates the frequency of co-citations of articles.

3. Antecedents of Smart Wearable Use

Prior research has identified multiple forces shaping smart wearable adoption. Early work emphasized technical performance, but recent studies reveal that sustained engagement depends on motivational, psychological, functional, and contextual factors working together. In this review, "mechanism" refers to the mediating processes through which wearable features and user perceptions translate into behavioral and wellbeing outcomes. Figure 3 illustrates the integrated AMC framework that organizes the key antecedents, mediating processes, and multilevel outcomes of smart wearable device use.

Motivational forces initiate adoption. Users are driven by hedonic enjoyment, health goals, interactive engagement, and self-expression. Hedonic motivation strengthens trust and usage intention[1], while health motivation sustains continued use[2]. Smartwatches serve as fashion-technology hybrids that enable identity expression and uniqueness seeking [3]. These motivations enhance perceived relevance and experiential value, triggering initial engagement.

Psychological need fulfillment sustains wearable use. Autonomy improves long-term adherence, while data-driven feedback and personalized insights support competence and relatedness [4]. Perceived health risks can heighten sensitivity to functional value, reinforcing usage intention [5]. Aligned with Self-Determination Theory, these findings suggest wearables operate through feedback-enabled self-regulation and need satisfaction within the AMC mechanism.

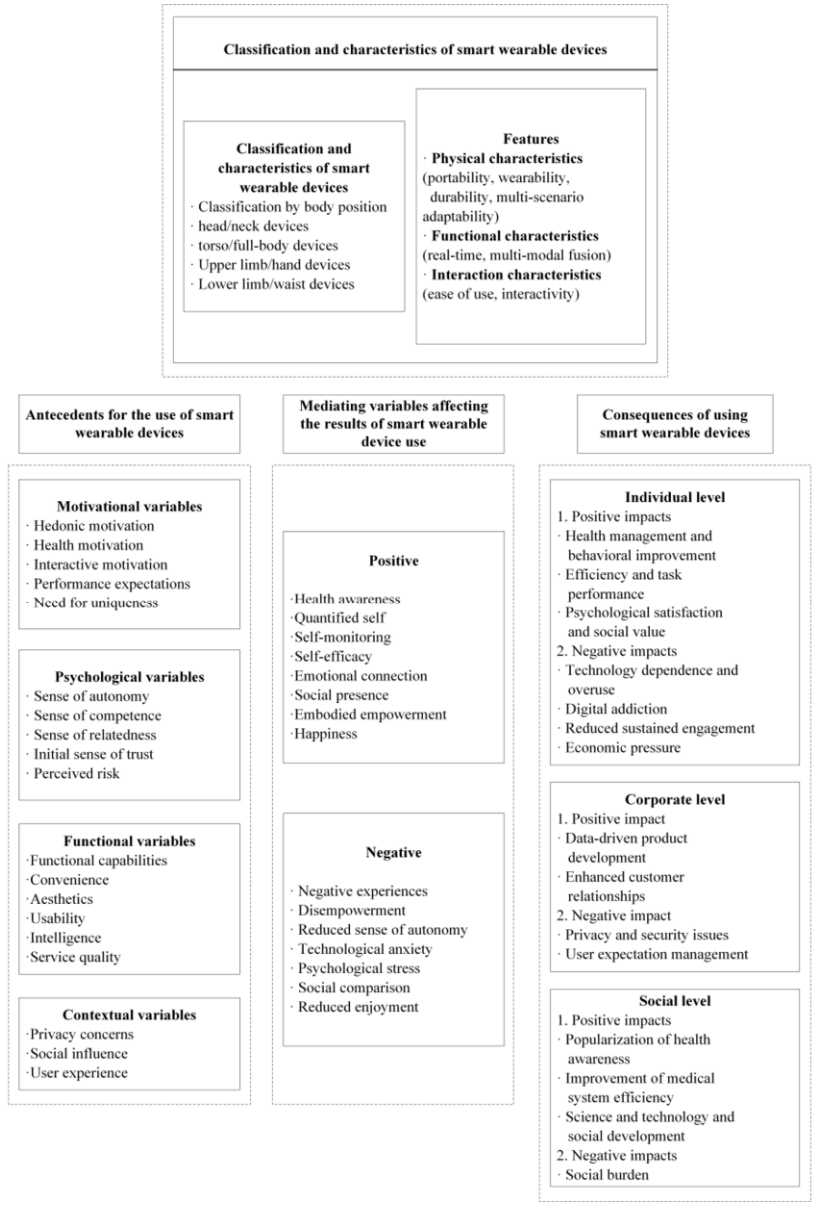


Figure 3 Smart Wearable Research Framework

Functional attributes enable adoption. Capability, usability, convenience, aesthetics, intelligence, and service quality shape performance expectations and ease of use. Real-time monitoring and personalized guidance improve health perceptions and social integration, especially among older adults [6]. Poor compatibility or insufficient support undermines trust and slows adoption [1]. Service quality, reflected in reliability, data accuracy, and user-friendliness, sustains engagement by strengthening empowerment [7]. Functional factors reduce friction and reinforce credibility.

Contextual conditions moderate adoption. Privacy concerns remain a barrier [1], while social influence, including norms and community participation, positively affects acceptance [8]. Experience-centered design, particularly gamification and tracking features, supports habit formation [9]. These variables shape the trust environment and social reinforcement surrounding use.

Wearable adoption reflects an integrated process: motivational relevance triggers engagement; psychological needs and feedback sustain involvement; functional performance enables use; and contextual conditions reinforce or constrain adoption. The AMC framework clarifies how wearable technologies translate perceptions into behavior.

4. Consequences of Using Smart Wearable Devices

Smart wearables generate multi-level consequences that are both enabling and constraining. Their effects follow a dual-edged pattern shaped by engagement intensity, technological reliability, and context. Prior work reveals distinct yet interconnected impacts at individual, organizational, and societal levels.

At the individual level, wearables enhance self-monitoring, health awareness, and behavioral regulation. Visualized feedback, such as sleep analytics and goal reminders, improves data interpretability and supports health behavior compliance [8]. Real-time monitoring and personalized feedback facilitate habit formation and sustained awareness [10]. Physiological benefits include improved sleep regulation [11] and performance enhancement via haptic feedback [12]. AI-enabled interactions strengthen satisfaction and identity reinforcement [10].

However, intensive self-tracking risks technological dependence and motivational crowding-out. Continuous monitoring has been linked to data intelligence discipline and digital addiction, particularly under high surveillance intensity [2]. Disengagement occurs when users experience technical fatigue, emotional burden, or declining social incentives. Excessive quantification may reframe leisure as obligation, weakening intrinsic motivation over time. The same feedback mechanisms that enable self-regulation may, under certain conditions, undermine autonomous engagement.

At the organizational level, wearable data infrastructures enable personalization and innovation. Firms leverage physiological and behavioral data for predictive modeling, product iteration, and engagement optimization [13]. Quantified interaction data strengthens customer relationship management by making user behavior observable and actionable [10]. These benefits emerge when data quality, system integration, and service responsiveness align.

However, business value is constrained by privacy sensitivity and expectation management risks. Prior research emphasizes balancing experience enhancement with data protection, as inflated expectations or perceived surveillance erode trust and loyalty. Organizational gains from wearables depend on maintaining a credible trust-value equilibrium.

At the societal level, wearable diffusion contributes to rising health consciousness and healthcare efficiency. Short-term adoption internalizes health norms and promotes preventive awareness [8], while system-level implementations demonstrate efficiency gains in healthcare delivery [14]. Wearables function as supportive tools within digital health ecosystems.

Yet societal risks emerge. Without appropriate governance, technological advancement may privilege commercial interests over public value, and pervasive self-tracking can intensify social comparison and role-related anxiety [15]. Wearable technologies amplify existing social dynamics rather than serving as neutral health tools.

The consequences of smart wearable use exhibit patterned ambivalence. Individual outcomes reflect a self-regulation–dependence tension; organizational outcomes hinge on a personalization–privacy balance; and societal outcomes reveal an efficiency–surveillance trade-off. The AMC framework shows that wearables create value when feedback, functionality, and trust conditions align. Future research should move beyond net-effect assumptions and examine boundary conditions under which wearable benefits can be sustained without triggering over-monitoring, privacy erosion, or motivational fatigue.

5. Summary and Research Agenda

This review systematically synthesized the conceptual foundations, antecedents, mechanisms, and consequences of smart wearable use across individual, organizational, and societal levels. While prior studies have generated valuable insights, the literature remains fragmented, particularly in explaining how antecedent conditions translate into outcomes through mediating processes.

To address this gap, the research agenda is structured around the Antecedent–Mechanism–Consequence (AMC) framework. Section 5.1 examines the dynamic and multidimensional nature of antecedents, Section 5.2 explores mediating mechanisms that connect drivers to outcomes, and Section 5.3 outlines long-term consequences across multiple levels. This agenda aims to move the field from factor identification toward mechanism clarification and longitudinal impact assessment, providing clearer directions for theory-driven future research.

5.1 Prospective Study: Systematic Research on Factors Influencing Smart Wearable Use

Users' adoption decisions reflect an interplay of health aspirations, privacy concerns, and social influences rather than single factors [2]. Motivation is dynamic, shifting from novelty and social influence in early adoption to reliability and accuracy in long-term use [16]. While social comparison can encourage engagement, it may also heighten anxiety [17]. Future research should treat motivations as multidimensional and evolving [18].

As wearables function within IoT ecosystems, interoperability and data integration have become critical [15]. Ecosystem interactivity and compatibility shape user engagement, yet ecosystem-level dynamics influencing retention and satisfaction remain underexplored. Wearable morphology now extends beyond watches to smart clothing, skin patches, and implantables [19], blurring distinctions between human and machine. Form influences acceptance, cognitive load, and perception of bodily boundarie. Future research should refine classification systems accounting for morphology, intimacy, and feedback modalities.

5.2 Research Agenda for Mechanism: Mediating Pathways of Wearable Use

Wearables influence not only health monitoring but also cognition, social interaction, and consumption. They enhance cognitive functioning in older adults through social interaction [20], reshape body awareness, and foster intrinsic motivation when data is shared—yet may reinforce surveillance and unrealistic ideals [21]. Cross-domain behavioral mediation remains underexplored.

Negative outcomes, reduced enjoyment, privacy concerns, and compulsive monitoring, require compensation strategies like human-centered data presentation [13] and well-being-focused design.

Device reliance risks eroding sensory trust: poor sleep reports can induce fatigue regardless of actual rest, and conflicting feedback increases cognitive load [22]. Balancing digital metrics with embodied awareness is critical to prevent overdependence.

5.3 Research Agenda for Consequence Variables: Long-Term Impacts of Wearable Use

At the individual level, continuous monitoring enhances health awareness and well-being [8], yet may cause data overload and obsessive self-tracking. Research should clarify how to optimize psychological benefits while mitigating anxiety, information fatigue, and pathological outcomes.

At the corporate level, rapid device iteration raises sustainability challenges across production and disposal lifecycles [23]. Modular design offers solutions to reduce waste, while digital transformation drives subscription-based, service-oriented models. Firms must integrate intelligence, style, and sustainability to align consumer and environmental needs.

At the societal level, wearables enhance healthcare efficiency but risk widening the digital divide. Adoption remains unequal across demographics—elderly, children, and low-income groups are underrepresented. Without equity-focused strategies, innovations may reinforce inequalities [24]. Interventions ensuring inclusive access, literacy, and democratized health data are critical.

References

[1] Gu Z, Xu F, Wei J. An Empirical Study on the Factors Influencing Initial Trust in Wearable Business Consumers[J/OL]. Management Review, 2015, 27(7): 168-176. DOI: 10.14120/j.cnki.cn11-5057/f.2015.07.017.

[2] Xu X, Zhao L. Research Progress and Prospects in Health Motivation[J]. Advances in Psychological Science, 2010, 18(9): 1496-1502.

[3] Choi J, Kim S. Is the smartwatch an IT product or a fashion product? A study on factors affecting the intention to use smartwatches[J/OL]. Computers in Human Behavior, 2016, 63: 777-786. DOI: 10.1016/j.chb.2016.06.007.

[4] Karapanos E, Gouveia R, Hassenzahl M, et al. Wellbeing in the Making: Peoples' Experiences with Wearable Activity Trackers[J/OL]. Psychology of Well-Being, 2016, 6(1): 4. DOI: 10.1186/s13612-016-0042-6.

[5] Khoa B T, Huynh T T. Why do generation X customers use wearable fitness technology equipment after recovering from coronavirus? The role of perceived health risks[J/OL]. Heliyon, 2024, 10(12): e32978. DOI: 10.1016/j.heliyon.2024.e32978.

[6] Stragier J, Vanden Abeele M, Mechant P, et al. Understanding persistence in the use of Online Fitness Communities: Comparing novice and experienced users[J/OL]. Computers in Human Behavior, 2016, 64: 34-42. DOI: 10.1016/j.chb.2016.06.013.

[7] Nelson E C, Verhagen T, Noordzij M L. Health empowerment through activity trackers: An empirical smart wristband study[J/OL]. Computers in Human Behavior, 2016, 62: 364-374. DOI: 10.1016/j.chb.2016.03.065.
[8] Stiglbauer B, Weber S, Batinic B. Does your health really benefit from using a self-tracking device? Evidence from a longitudinal randomized control trial[J/OL]. Computers in Human Behavior, 2019, 94: 131-139. DOI: 10.1016/j.chb.2019.01.018.
[9] Oc Y, Plangger K. GIST do it! How motivational mechanisms help wearable users develop healthy habits[J/OL]. Computers in Human Behavior, 2022, 128: 107089. DOI: 10.1016/j.chb.2021.107089.
[10] Puntoni S, Reczek R W, Giesler M, et al. Consumers and Artificial Intelligence: An Experiential Perspective[J/OL]. Journal of Marketing, 2021, 85(1): 131-151. DOI: 10.1177/0022242920953847.
[11] Guarana C L, Barnes C M, Ong W J. The effects of blue-light filtration on sleep and work outcomes.[J/OL]. Journal of Applied Psychology, 2021, 106(5): 784-796. DOI: 10.1037/apl0000806.
[12] Hadi R, Valenzuela A. Good Vibrations: Consumer Responses to Technology-Mediated Haptic Feedback[J/OL]. Journal of Consumer Research, 2020, 47(2): 256-271. DOI: 10.1093/jcr/ucz039.
[13] King D, Auschaitrakul S, You Y. Felt something, hence it works: Merely adding a sensory signal to a product improves objective measures of product efficacy and product evaluations[J/OL]. Journal of the Academy of Marketing Science, 2024, 52(6): 1761-1779. DOI: 10.1007/s11747-024-01030-z.
[14] Băjenaru L, Marinescu I A, Dobre C, et al. Identifying the Needs of Older People for Personalized Assistive Solutions in Romanian Healthcare System[J/OL]. Studies in Informatics and Control, 2020, 29(3): 363-372. DOI: 10.24846/v29i3y202009.
[15] Reda R, Piccinini F, Martinelli G, et al. Heterogeneous self-tracked health and fitness data integration and sharing according to a linked open data approach[J/OL]. Computing, 2022, 104(4): 835-857. DOI: 10.1007/s00607-021-00988-w.
[16] El-Gayar O, Elnoshokaty A. Factors and Design Features Influencing the Continued Use of Wearable Devices[J/OL]. Journal of Healthcare Informatics Research, 2023, 7(3): 359-385. DOI: 10.1007/s41666-023-00135-4.
[17] Pennington N, Dam L. Social interaction, support preferences, and the use of wearable health trackers[J/OL]. Frontiers in Communication, 2023, 8: 1256452. DOI: 10.3389/fcomm.2023.1256452.
[18] Pearl D, Intriligator J, Liu X. Seamless Integration: The Evolution, Design, and Future Impact of Wearable Technology[A/OL]. arXiv, 2025. DOI: 10.48550/arXiv.2502.05797.
[19] Nature. Poke, twist, vibrate: a wearable device that delivers touch to the skin[J/OL]. Nature, 2025: d41586-025-00340-00348. DOI: 10.1038/d41586-025-00340-8.
[20] Cai T, Ma S, Zhong R. Longitudinal relationships between smart wearables, social interaction, and cognitive function: A cross-lagged panel model[J/OL]. Geriatric Nursing, 2024, 60: 399-407. DOI: 10.1016/j.gerinurse.2024.10.008.
[21] Berry R A, Rodgers R F, Campagna J. Outperforming iBodies: A Conceptual Framework Integrating Body Performance Self-Tracking Technologies with Body Image and Eating Concerns[J/OL]. Sex Roles, 2021, 85(1-2): 1-12. DOI: 10.1007/s11199-020-01201-6.
[22] Morris C, Danry V, Maes P. Wearable systems without experiential disruptions: exploring the impact of device feedback changes on explicit awareness, physiological synchrony, sense of agency, and device-body ownership[J/OL]. Frontiers in Computer Science, 2023, 5: 1289869. DOI: 10.3389/fcomp.2023.1289869.
[23] Gurova O, Merritt T R, Papachristos E, et al. Sustainable Solutions for Wearable Technologies: Mapping the Product Development Life Cycle[J/OL]. Sustainability, 2020, 12(20): 8444. DOI: 10.3390/su12208444.
[24] Badr J, Motulsky A, Denis J L. Digital health technologies and inequalities: A scoping review of potential impacts and policy recommendations[J/OL]. Health Policy, 2024, 146: 105122. DOI: 10.1016/j.healthpol.2024.105122.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261072

A Study on the Influencing Factors of Tacit Knowledge Sharing Intention among Female Hotel Managers: A TPB-Based Perspective

Yidan Zhu[a,1], Yingjie Liu[b,2] and Yuqian Liu[a,3]
[a] *Hunan Women's University, China*
[b] *Hunan Normal University, China*
ORCiD ID: Yidan Zhu https://orcid.org/0009-0006-4583-801X
Yingjie Liu https://orcid.org/0009-0003-3081-2341

Abstract. Grounded in the Theory of Planned Behavior (TPB), this study investigates the mechanisms through which workplace loneliness and perceived over-qualification influence the tacit knowledge sharing intention of female hotel managers. Self-efficacy is introduced as a mediating variable to empirically test the proposed model. Based on a survey conducted among female hotel managers, 138 valid responses were collected and analyzed using structural equation modeling. The results indicate that workplace loneliness has a significant negative impact on tacit knowledge sharing intention, while perceived over-qualification influences the intention indirectly through self-efficacy. Self-efficacy plays a mediating role in both relationships. The model demonstrates good fit indices (SRMR=0.035, NFI=0.872). This study extends the application of TPB in the field of knowledge management and offers theoretical and practical implications for designing incentive mechanisms and organizational support systems tailored to female managers in the hotel industry.

Keywords. Tacit knowledge sharing intention, female managers, workplace loneliness, perceived over-qualification, self-efficacy

1. Introduction

In the context of globalization and rapid digital transformation, women are playing an increasingly prominent role in the workforce. This trend is particularly evident in people-oriented service industries such as the hotel sector, where female hotel managers constitute a substantial proportion of the managerial workforce and exert significant

[1]First author: Yidan Zhu, Lecturer, Ph.D., research interests in Hotel and Homestay Digital Intelligent Operation; E-mail: zhuyidan@hnwu.edu.cn
[2] Corresponding Author and Second author: Yingjie Liu, Professor and master's supervisor of the School of Tourism, Hunan Normal University, email: liuyj@hunnu.edu.cn
[3]Third author: Yuqian Liu; the student at Hunan women's university in major of hospitality Management E-mail: m19375102120@163.com
Funded by: The research project on teaching reform of postgraduate education in Hunan Province, "Exploration and Practice of Integrated Undergraduate and Postgraduate Education for Tourism Application-oriented Talents in the Information Age" (2024JGZD032)

influence on managerial decision-making, knowledge management, and organizational innovation. According to the Outline for the Development of Chinese Women (2021-2030), the number of employed women in China had reached 320 million by 2022, accounting for 43.2% of the total employed population, with women representing 40.5% [1] of employees in urban non-private units. A white paper released by the Hospitality Industry Pipeline Coalition reported that women account for approximately 70% of the global tourism and hospitality workforce, underscoring their growing importance in this industry.

Owing to personal attributes such as attentiveness, patience, and strong empathy, women often demonstrate unique advantages in hotel management contexts [2]. As consumer demand becomes increasingly diversified and industry competition intensifies, hotel management has gradually shifted from a focus on tangible assets to an emphasis on intangible capabilities, particularly knowledge management. Knowledge sharing, as a core component of organizational knowledge management, not only shapes individual job performance but also directly affects organizational innovation and competitiveness [3]. Against this backdrop, understanding how female hotel managers engage in tacit knowledge sharing under high-pressure working conditions has become a critical issue warranting scholarly attention.

Although women have increasingly occupied middle- and senior-level management positions in the hotel industry, they continue to face challenges stemming from organizational culture, gender stereotypes, and family responsibilities [4-6]. These pressures may suppress their motivation to engage in tacit knowledge sharing, which is inherently experiential, difficult to formalize, and highly dependent on psychological safety and interpersonal trust [7]. For female managers, the sharing of tacit knowledge thus represents not only a professional behavior but also a psychologically embedded decision shaped by emotional and cognitive experiences.

At present, systematic research on tacit knowledge sharing intention among female hotel managers remains limited. Existing studies have largely focused on women's career barriers, leadership differences, gender bias, or general knowledge management issues, while empirical research examining how emotional experiences and cognitive perceptions jointly shape tacit knowledge sharing intention in hotel management contexts remains scarce. Consequently, clear gaps persist regarding variable integration, theoretical explanation, and contextual specificity.

To address these gaps, this study adopts the Theory of Planned Behavior (TPB) as its theoretical foundation and focuses on female hotel managers as the research target. By integrating workplace loneliness and perceived over-qualification as emotional and cognitive antecedents and introducing self-efficacy as a mediating variable, this study constructs a theoretical model to explain the formation mechanism of tacit knowledge sharing intention. The findings are expected to extend the application of TPB in knowledge management and gender-related research and provide insights for hotel organizations seeking to foster supportive environments for female managers.

2. Literature Review and Research Hypothesis Development

2.1. Theory of Planned Behavior and Tacit Knowledge Sharing Intention

Tacit knowledge sharing intention refers to an individual's subjective willingness to transfer experiential, skill-based, and intuition-driven knowledge that is difficult to

codify or formalize within an organization. As a critical form of internal knowledge flow, tacit knowledge sharing plays a vital role in promoting organizational learning, improving performance, and enhancing innovation capability. However, due to the implicit and personal nature of tacit knowledge, its sharing process is particularly sensitive to individual psychological states and organizational contexts [8].

Prior studies have examined the antecedents of tacit knowledge sharing from multiple perspectives, including organizational climate, leadership trust, personality traits, and employee emotions. A fair and reciprocal organizational culture and managerial support have been shown to enhance knowledge sharing intention [9], while individual characteristics such as self-efficacy and extraversion are positively associated with knowledge sharing behaviors [10]. These findings indicate that tacit knowledge sharing is not only a rational exchange behavior, but also a psychologically embedded behavioral intention.

The Theory of Planned Behavior (TPB) provides a useful framework for explaining behavioral intentions. According to TPB, intention is shaped by attitude toward the behavior, subjective norms, and perceived behavioral control, with perceived behavioral control reflecting individuals' assessment of their capability and resources to perform a behavior [11]. Self-efficacy is closely related to perceived behavioral control because it captures individuals' confidence in successfully performing knowledge sharing behaviors.

Recent TPB-based research has incorporated emotional and cognitive antecedents into the explanation of behavioral intention. Emotional factors such as trust and psychological safety influence behavioral attitudes, whereas cognitive factors such as self-efficacy and identity perceptions shape perceived behavioral control [10][12]. Workplace loneliness may indirectly constrain knowledge sharing intention by weakening self-efficacy [13], while perceived over-qualification may affect knowledge sharing through perceived behavioral control and occupational identity threats [14]. Therefore, this study adopts TPB as its theoretical foundation and introduces self-efficacy as a mediating mechanism linking workplace loneliness and perceived over-qualification to tacit knowledge sharing intention among female hotel managers.

2.2. Workplace Loneliness and Self-Efficacy

Workplace loneliness refers to employees' perceived lack of meaningful interpersonal connection and unmet belongingness needs in the workplace [12]. Prior studies indicate that workplace loneliness is associated with weaker social connection, lower workplace adjustment, and reduced help or knowledge-related behaviors [8][9]. From a psychological resource perspective, lonely employees may perceive lower organizational support and interpersonal trust, which can weaken their confidence in handling tasks and communicating effectively. This mechanism is particularly relevant for female hotel managers, who may face gender stereotypes, limited informal networks, and constrained leadership opportunities [2] [4-6]. When workplace loneliness reduces perceived support and psychological security, female managers may become less confident in solving problems, coordinating with colleagues, and contributing expertise. Prior research further suggests that self-efficacy is an important psychological mechanism linking workplace loneliness to tacit knowledge sharing intention [13]. Therefore, the following hypothesis is proposed:

H1: Workplace loneliness has a significant effect on self-efficacy.

2.3. Perceived Over-qualification and Self-Efficacy

Perceived over-qualification denotes employees' perception that their education, skills, experience, or abilities exceed job requirements. Prior studies show that perceived over-qualification may lead to negative attitudes and behaviors, including lower engagement, alienation, and knowledge hiding, when surplus competence is interpreted as person-job misfit [7][14][15]. However, it also reflects competence-based self-evaluation. Employees who perceive themselves as over-qualified may recognize that they possess knowledge and abilities beyond current job demands, thereby strengthening their confidence in handling work tasks and contributing expertise. In knowledge-sharing contexts, this perception may either trigger resource-protection motives or promote contributions of knowledge, depending on how employees evaluate their surplus capability [16][17]. Therefore, perceived over-qualification may shape female hotel managers' self-efficacy by influencing their perceived professional competence.

H2: Perceived over-qualification has a significant effect on self-efficacy.

2.4. The Effect of Self-Efficacy on Tacit Knowledge Sharing Intention

Self-efficacy, as defined in social cognitive theory, refers to individuals' beliefs in their ability to successfully perform specific tasks and achieve desired outcomes. Within the TPB framework, self-efficacy represents a core component of perceived behavioral control and plays a pivotal role in shaping behavioral intentions.

Empirical evidence in the knowledge management literature indicates that self-efficacy is a strong positive predictor of knowledge sharing intention [8][10]. Employees with high self-efficacy are more confident in the value of their knowledge and more willing to share their expertise to gain recognition and exert influence within the organization. Employees with higher levels of self-efficacy are more confident in the value of their knowledge and tend to achieve influence and recognition through sharing. Self-efficacy has a significant positive effect on tacit knowledge sharing intention, and particularly when individuals face emotional barriers such as loneliness, self-efficacy plays a buffering and motivating role [12].

H3: Self-efficacy has a significant effect on tacit knowledge sharing intention.

2.5. The Mediating Role of Self-Efficacy

Existing studies highlight self-efficacy as an important mediator linking negative psychological states and behavioral intention. Prior research confirms that self-efficacy partially mediates the relationship between workplace loneliness and tacit knowledge sharing intention [12] and significantly mediates the effect of perceived over-qualification on knowledge sharing behavior, particularly in collectivist cultural contexts [13]. As a cognitive mechanism, self-efficacy translates emotional and cognitive experiences into behavioral intention.

H4: Self-efficacy mediates the relationship between workplace loneliness and tacit knowledge sharing intention.

H5: Self-efficacy mediates the relationship between perceived over-qualification and tacit knowledge sharing intention.

Grounded in the Theory of Planned Behavior proposed by Ajzen (1991) [11], this study examines how workplace loneliness and perceived over-qualification influence

tacit knowledge sharing intention among female hotel managers, with self-efficacy serving as a mediating variable. The conceptual research model is illustrated in Figure 1.

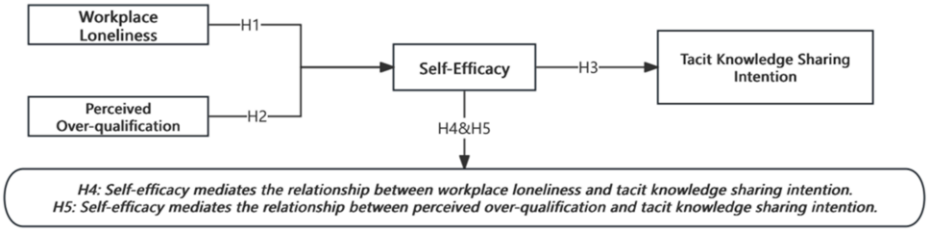


Figure 1. Conceptual Model of Tacit Knowledge Sharing Intention of Female Hotel Managers

3. Measurement of Variables

3.1. Workplace Loneliness

Workplace loneliness is defined as individuals' perceived social isolation and emotional detachment resulting from insufficient interpersonal connections and a lack of belonging within the organization (see Table 1). This study adopts the Chinese version of the workplace loneliness scale revised by Ma (2022) based on the Loneliness at Work Scale (LAWS), developed by empirically validated in the Chinese organizational context [12].

Table 1. Measurement Items of Workplace Loneliness

Number	Item	References
WL1	I feel that I am unimportant and lack indispensable value in the organization.	Ma,2022
WL2	I feel that I have not received recognition or appreciation from my supervisors or colleagues within the organization.	
WL3	I often feel a sense of powerlessness at work and tend to leave the workplace as early as possible after finishing my workday.	
WL4	My supervisors and colleagues seem indifferent to my long-term contributions, making me feel that I am insignificant.	
WL5	I find my current job monotonous and boring, and I am unable to derive any sense of meaning or value from it.	
WL6	During my work, I often perceive a sense of emotional distance from my colleagues.	
WL7	I feel that I am unable to truly integrate with my colleagues as part of a team or collective.	
WL8	I believe that my colleagues are often emotionally cold or distant toward me.	
WL9	When I encounter difficulties or challenges at work, I do not confide in my colleagues.	
WL10	I perceive a competitive relationship with my colleagues, and I do not believe we can become genuine friends.	

3.2. Perceived Over-qualification

Perceived over-qualification (POQ) refers to individuals' perceptions that their education, experience, or skills exceed the requirements of their current position (see Table 2). This study drew upon the "Subjective Perception of Extra-Personal Abilities Assessment Scale" (SPOQ) that Zhao (2023) [13] and Zhu (2023) [14] validated in their Chinese employee samples.

Table 2. Measurement Items of Perceived Over-qualification

Number	Item	References
POQ1	My job requires less education than I possess.	
POQ2	The work experience I have is not necessary for successfully performing my job.	
POQ3	I possess job skills that are not required for my current position.	
POQ4	Someone with a lower level of education than mine could perform my job well.	Zhao,2023;
POQ5	My previous training has not been fully utilized in my current job.	Zhu,2023
POQ6	To perform my job, I have a great deal of knowledge that is not needed.	
POQ7	My level of education exceeds the educational requirements of my job.	
POQ8	Someone with less work experience than I could be competent in my job.	
POQ9	I have more abilities than are required to perform my job.	

3.3. Self-Efficacy

Self-efficacy refers to individuals' beliefs in their ability to cope with challenges, accomplish tasks, and achieve goals, and it is one of the core variables representing perceived behavioral control in the Theory of Planned Behavior. A high level of self-efficacy can effectively enhance employees' positive behavioral intentions, including communication, cooperation, and knowledge sharing. This study employed the Self-Efficacy Scale (GSES) developed by Ma (2022) for the field of organizational behavior and management research in China [12]. This scale consists of 10 items (see Table 3).

Table 3. Measurement Items of Self-Efficacy

Number	Item	References
SE1	If I put in the necessary effort, I can always solve difficult problems.	
SE2	Even when others oppose me, I can usually find ways to obtain what I need.	
SE3	It is easy for me to persist in my goals and achieve them.	
SE4	I am confident that I can calmly and effectively handle unexpected situations.	
SE5	With my personal abilities, I can deal well with unforeseen circumstances.	
SE6	If I make the necessary effort, I can solve most problems.	Ma,2022
SE7	I can remain calm when facing difficulties because I am confident in my ability to handle them.	
SE8	When confronted with a problem, I can usually think of several solutions.	
SE9	When I encounter difficulties, I am generally able to find ways to resolve them.	
SE10	No matter what happens to me, I can cope with it effectively.	

3.4. Tacit Knowledge Sharing Intention

Tacit knowledge sharing intention refers to individuals' willingness to actively share experiential, skill-based, and insight-driven knowledge that is difficult to codify within the organization. This measurement tool adopts the empirically validated version by Ma (2022) [12] and includes six items applicable to the organizational environment in China (see Table 4).

Table 4. Measurement Items of Tacit Knowledge Sharing Intention

Number	Item	Reference
SS1	I am willing to share my work experiences with my colleagues.	
SS2	If my colleagues need and request it, I am willing to share my specialized professional skills.	
SS3	I am willing to provide my personal work notes and other learning materials to my colleagues.	Ma,2022
SS4	I am willing to guide and mentor my colleagues through work-related knowledge sharing.	

SS5	I am willing to share my personal experiences, ideas, and work-related skills with my colleagues.
SS6	I am willing to share creative ideas and inspirations with my colleagues.

4. Data Analysis

4.1. Sample Characteristics

Given the research focus on female hotel managers, the target population of this study comprises female managerial professionals working in the hotel industry. Specifically, the sample includes female managers currently employed or previously employed in high-end international brand hotels in China. To enhance the representativeness and relevance of the sample, respondents were drawn from internationally branded hotels located in Shanghai and Changsha, covering middle- and senior-level management positions in departments such as front office, housekeeping, human resources, and marketing.

Data was collected between February and March 2025 using a combination of online and offline surveys. A total of 171 questionnaires were returned. After excluding 33 responses completed by male participants, 138 valid questionnaires were retained for analysis, yielding an effective response rate of 80.70%.

Regarding age, 14.49% of respondents were aged 18–24 (n = 20), 36.23% were aged 25–40 (n = 50), 39.86% were aged 41–60 (n = 55), and 9.42% were aged 61 and above (n = 13). No respondents were under 18 years old.

In terms of education, 49.28% of respondents held a junior college degree (n = 68) and 37.68% held a bachelor's degree (n = 52), together accounting for 87.0% of the sample. Respondents with high school or technical secondary education accounted for 9.42% (n = 13), while those with postgraduate degrees or below junior high school level accounted for less than 4%.

4.2. Reliability and Validity Analysis

SmartPLS 4.0 was employed to assess the reliability and validity of the measurement model. As reported in Table 5, Cronbach's alpha, rho_A, and composite reliability (CR) values for all constructs exceeded 0.70, indicating satisfactory internal consistency. In addition, the average variance extracted (AVE) values ranged from 0.709 to 0.747, surpassing the recommended cutoff of 0.50 and confirming adequate convergent validity. These results suggest that the measurement model is reliable and suitable for structural model estimation.

Table 5. Reliability and Validity Statistics

Construct	Cronbach's alpha	rho_A	CR	AVE
Workplace Loneliness	0.954	0.954	0.960	0.709
Self-Efficacy	0.962	0.962	0.967	0.745
Perceived Over-qualification	0.953	0.953	0.960	0.727
Tacit Knowledge Sharing Intention	0.932	0.934	0.946	0.747

4.3. Structural Equation Modeling Analysis

The structural model was estimated using SmartPLS 4.0, and the significance of the hypothesised paths was assessed through the bootstrapping procedure. Structural

equation modeling (SEM) was employed to examine the hypothesised relationships among workplace loneliness, perceived over-qualification, self-efficacy, and tacit knowledge sharing intention. Model fit was first assessed to ensure the adequacy of the structural model (Table 6). As reported in Table 8, the overall model exhibited an acceptable fit to the data. The standardized root mean square residual (SRMR) was 0.035, indicating a good approximation between the observed and model-implied correlation matrices. In addition, the normed fit index (NFI) reached 0.872, suggesting a satisfactory level of explanatory power for the estimated model.

Table 6. Model Fit Indices

Fit Index	Saturated model	Estimated model
SRMR	0.035	0.035
d_ULS	0.756	0.756
d_G	0.932	0.932
Chi-square	817.351	817.351
NFI	0.872	0.872

The original SmartPLS output for the direct paths is reported in Table 7. The results show that perceived over-qualification had a significant positive effect on self-efficacy, indicating that female hotel managers who perceived themselves as over-qualified tended to report stronger confidence in their professional competence. Workplace loneliness also had a significant positive effect on self-efficacy, suggesting that workplace loneliness was significantly associated with respondents' perceived capability in work-related situations. In addition, self-efficacy exerted a significant positive effect on tacit knowledge sharing intention, confirming that efficacy beliefs played an important role in promoting willingness to share experiential and implicit knowledge. These results support H1, H2, and H3 and provide the statistical basis for further testing the mediating role of self-efficacy.

Table 7. Structural Model Path Coefficients

Path coefficients	Original sample (O)	Sample mean (M)	Standard deviation	T statistics	P values
H1:Workplace Loneliness →Self-efficance	0.488	0.488	0.061	8.034	0.000
H2:Perceived Over qualification → Self-efficance	0.423	0.424	0.064	6.592	0.000
H3:Self-efficance →Tacit Knowledge Sharing Intention	0.425	0.423	0.094	4.536	0.000

4.4. Mediation Effect Analysis

To further examine the mediating mechanism, the specific indirect effects were tested using the bootstrapping procedure in SmartPLS 4.0, and the results are summarised in Table 8. The indirect effect of workplace loneliness on tacit knowledge sharing intention through self-efficacy was positive and statistically significant, supporting H4. Similarly, the indirect effect of perceived over-qualification on tacit knowledge sharing intention through self-efficacy was also significant, supporting H5. These findings indicate that self-efficacy serves as a mediating mechanism through which workplace loneliness and perceived over-qualification influence tacit knowledge sharing intention. Since no interaction term was specified or tested in the model, the results should be interpreted as evidence of mediation rather than moderation.

Table 8. Mediation Effects

Specific indirect effects	Original sample (O)	Sample mean (M)	Standard deviation	T statistics	P values
H4:Workplace Loneliness →Self-efficance→Tacit Knowledge Sharing Intention	0.208	0.206	0.052	4.003	0.000
H5:Perceived Over-qualification→Self-efficance→Tacit Knowledge Sharing Intention	0.180	0.179	0.048	3.722	0.000

5. Conclusions and Implications

5.1. Discussion

Grounded in the Theory of Planned Behavior (TPB), this study elucidates how emotional and cognitive experiences shape tacit knowledge sharing intention among female hotel managers through self-efficacy. The findings suggest that tacit knowledge sharing is a psychologically embedded process rather than a purely rational decision.

Workplace loneliness emerges as a salient emotional antecedent of tacit knowledge sharing intention. Consistent with prior research, workplace loneliness undermines interpersonal trust, psychological safety, and the sense of belonging, thereby inhibiting discretionary behaviors such as knowledge sharing [3][8][16]. In hotel organizations characterized by intensive emotional labor, tacit knowledge exchange relies heavily on informal interaction and mutual trust. For female hotel managers, workplace loneliness may be further reinforced by gendered organizational structures and limited access to informal managerial networks [4-6], increasing emotional withdrawal and reducing willingness to share experiential knowledge. From a TPB perspective, workplace loneliness weakens behavioral attitudes and perceived social support, constraining knowledge sharing intention.

The effect of perceived over-qualification is more nuanced. Although perceived over-qualification does not directly affect tacit knowledge sharing intention, it significantly enhances self-efficacy, which in turn promotes knowledge sharing. This finding challenges the view that perceived over-qualification uniformly leads to negative outcomes such as withdrawal or knowledge hiding [7][8][17]. Instead, it indicates that perceived over-qualification may generate competence-based confidence that only translates into knowledge sharing when activated through perceived behavioral control. This pattern aligns with recent evidence emphasizing the conditional effects of perceived over-qualification on knowledge-related behaviors [15][17][18].

Most importantly, self-efficacy functions as the central psychological mechanism linking both workplace loneliness and perceived over-qualification to tacit knowledge sharing intention. Consistent with TPB, self-efficacy enhances individuals' confidence in their ability to share valuable knowledge and to exert influence through such behavior [13]. By integrating emotional and cognitive antecedents into a unified framework, this study extends existing knowledge-sharing research and clarifies the mechanisms underlying female managers' tacit knowledge sharing intention in hotel contexts.

5.2. Conclusions

This study examined the formation mechanism of tacit knowledge sharing intention among female hotel managers by integrating workplace loneliness and perceived over-qualification within a Theory of Planned Behavior (TPB) framework and introducing self-efficacy as a mediating variable. The findings demonstrate that workplace loneliness suppresses tacit knowledge sharing intention both directly and indirectly through reduced self-efficacy, whereas perceived over-qualification influences knowledge sharing exclusively via self-efficacy. These results highlight self-efficacy as a pivotal psychological mechanism shaping female managers' willingness to share tacit knowledge.

Theoretically, this study extends TPB by incorporating emotional and cognitive antecedents into the explanation of tacit knowledge sharing intention in a gender-specific managerial context. It advances tourism and hospitality research by moving beyond generic employee samples and clarifying the psychological pathways through which female managers' workplace experiences translate into knowledge-sharing intentions.

Practically, the findings suggest that hotel organizations should address emotional isolation and enhance self-efficacy among female managers through inclusive leadership practices and recognition of professional competence. Several limitations should be noted, including the cross-sectional design and context-specific sample. Future research may employ longitudinal designs or comparative approaches to further refine the proposed framework.

5.3. Managerial Implications

This study offers several concise managerial implications for hotel organizations seeking to promote tacit knowledge sharing among female managers. Hotel organizations should recognize workplace loneliness as a critical yet often overlooked barrier to knowledge sharing. Beyond formal structures, managers should foster inclusive leadership practices and informal managerial networks that enhance social embeddedness and psychological safety, thereby creating conditions conducive to experiential knowledge exchange.

Perceived over-qualification among female managers should not be treated solely as a risk factor. When appropriately recognized and leveraged, over-qualification can strengthen self-efficacy and indirectly stimulate tacit knowledge sharing. Hotel organizations may benefit from aligning managerial roles with individual competencies and providing opportunities for professional recognition, thus transforming perceived surplus capability into a resource for organizational learning.

The findings underscore self-efficacy as a key psychological lever. Managerial interventions that reinforce female managers' confidence, such as empowerment-oriented leadership, autonomy in decision-making, and feedback emphasizing expertise, are likely to be more effective than purely incentive-based knowledge management systems in fostering sustainable tacit knowledge sharing.

References

[1] National Working Committee on Children and Women under the State Council of China. Outline for the Development of Chinese Women (2021−2030) [Internet]. Beijing: National Working Committee on Children and Women under the State Council of China; 2021 [cited 2021 Sep 27]. Available from: http://www.nwccw.gov.cn/

[2] Xiong W, Chen S, Okumus B, et al. Gender stereotyping and its impact on perceived emotional leadership in the hospitality industry: a mixed-methods study. Tourism Management. 2022; 90:104476.
[3] Bryan BT, Andrews G, Thompson KN, et al. Loneliness in the workplace: a mixed-method systematic review and meta-analysis. Occupational Medicine. 2023;73:557−567.
[4] Kumara YADR. Career development of women in the hotel industry: an overview. Journal of Applied and Natural Science. 2018;10(1):330−338.
[5] Li T. Barriers to the development of female leadership and pathways for improvement from the perspective of leadership power sources. Manager Journal. 2024; (12):78−79.
[6] Li Y, Chen J. An exploratory study on the barriers to continuing career development of female middle managers. Journal of Tianjin University (Social Sciences Edition). 2024;26(4):308−319.
[7] Erdogan B, Karaeminogullari A, Bauer TN, et al. Perceived over qualification at work: implications for extra-role behaviors and advice network centrality. Journal of Management. 2020.
[8] Du S, Ma Y, Lee JY. Workplace loneliness and the need to belong in the era of COVID-19. Sustainability. 2022;14(8):4788.
[9] Zheng C, Zhou J, Liu S, et al. The impact of workplace loneliness on employees' helping behavior: a conservation of resources theory perspective. Journal of Hunan University of Technology (Social Sciences Edition). 2022;27(1):43−51.
[10] Zhao H, Li X. The mechanisms through which perceived overqualification affects organizational knowledge sharing behavior: the moderating role of traditional Chinese cultural values. Journal of Zhengzhou University (Philosophy and Social Sciences Edition). 2022;55(1):51−55.
[11] Ajzen I. The theory of planned behavior. Organizational Behavior and Human Decision Processes. 1991;50(2):179−211.
[12] Ma J. A study on the impact of workplace loneliness on employees' tacit knowledge sharing intention [master's thesis]. Beijing: Beijing Jiaotong University; 2022.
[13] Zhao J. A study on the impact of workplace loneliness on interpersonal helping behavior [master's thesis]. Beijing: University of International Business and Economics; 2022.
[14] Zhu Y. A study on the impact of hotel managers' perceived subordinate over qualification on knowledge hiding behavior [master's thesis]. Chengdu: Sichuan Agricultural University; 2024.
[15] Shang Z, Zuo C, Shi Y, et al. Does overqualification play a promoting or a hindering role? The impact of public employees' perceived overqualification on workplace behaviors. Behavioral Sciences. 2024;14(1):48.
[16] Zhao L, Zhao S, Zeng H, et al. To share or not to share? A moderated mediation model of the relationship between perceived overqualification and knowledge sharing. Baltic Journal of Management. 2021;16(4):533−555.
[17] Khan J, Saeed I, Zada M, et al. The positive side of overqualification: examining perceived overqualification linkage with knowledge sharing and career planning. Journal of Knowledge Management. 2022;26(6):1390−1412.
[18] Tomás I, González-Romá V, Valls V, et al. Perceived overqualification and work engagement: the moderating role of organizational size. Current Psychology. 2023;42:23069−23079.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)

© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261073

The Impact of Digital Transformation on Corporate Market Value: The Moderating Effect of Top Management Team Heterogeneity

Ziyang Liu[1], Shijie Bai and Jing Sun
Beijing Wuzi University, China
ORCiD ID: Ziyang Liu https://orcid.org/0009-0009-3165-5329

Abstract. Based on data from Chinese A-share listed companies from 2012 to 2022, this paper verifies the relationship between digital transformation and corporate market value. It further introduces top management team (TMT) background heterogeneity as a moderating variable to explore its regulatory effect on this relationship. The findings indicate that the degree of corporate digital transformation significantly promotes the generation of corporate market value. However, TMT heterogeneity inhibits the positive effect of digital transformation on market value. Specifically, among the various dimensions of TMT heterogeneity, financial background heterogeneity plays the primary role in exerting this inhibitory effect. The conclusions of this paper provide certain management implications for enterprises to enhance market value through digital transformation and offer suggestions for the construction of corporate TMT structures, holding theoretical and practical significance.

Keywords. Digital Transformation, Market Value, Top Management Team Heterogeneity

1. Introduction

1.1. Research Background

Digital transformation has become a core national strategy for systemic economic and societal restructuring, providing institutional support and directional impetus for the reinvention of firm value. The 14th Five-Year Plan mandates accelerated digital development and the construction of a digital economy, society, and government, thereby driving comprehensive change in production, lifestyles, and governance. This policy framing elevates digital transformation beyond technology adoption to a strategic force reshaping organizational structures and operational mechanisms. Enabled by new generation technologies such as cloud computing, big data, and the Internet of Things, firms can reconstruct customer experiences, streamline workflows, and overhaul business models, thereby enhancing productivity and reconfiguring value creation systems. A central scholarly concern, therefore, is how firms effectively convert digital strategy into tangible market value under evolving institutional guidance.

[1] Corresponding Author, Ziyang Liu, Beijing Wuzi University, China; Email: 670226003@qq.com

Although extant research broadly affirms a positive digital transformation–performance link, a marked divergence between macro trends and micro empirical findings points to underexplored moderating mechanisms. Prior studies show that digital transformation improves ESG performance (Song et al., 2024) [1], spurs innovation driven value creation (Wang et al., 2024) [2], strengthens indigenous value appropriation (Zhang et al., 2024) [3], and promotes market value growth via enhanced market value (Wang et al., 2023) [4], an effect amplified by peer behavior (Huang & Xie, 2022) [5]. Yet, the Report on the Development of China's Digital Economy Enterprises reveals that the average market capitalization of the Top 500 Chinese digital economy firms declined from RMB 154.7 billion in 2022 to RMB 87.2 billion in 2025. This contrast implies that the digitalization–market value pathway is not linear but contingent on contextual conditions, notably internal governance and strategic decision mechanisms, that moderate value realization. Identifying such moderators is thus essential, and the top management team (TMT) warrants focused attention as the locus of strategic choice.

TMT heterogeneity holds strong theoretical potential to moderate the digital transformation–market value nexus. Upper Echelons Theory posits that executives' cognitive bases and experiences shape strategic decisions, with heterogeneous teams enabling complementary knowledge and collaboration (Mason et al., 2004) [6]. In strategic emerging industries, age and tenure heterogeneity exhibit inverted U shaped effects on innovation efficiency, while educational and functional heterogeneity exert sustained positive impacts (Han et al., 2017) [7]. In digital contexts, TMT age heterogeneity correlates negatively with digital transformation progress, whereas heterogeneity in education, specialization, and overseas experience correlates positively (Tang et al., 2022) [8]. These findings indicate that TMT heterogeneity is not monolithic; distinct dimensions serve differentiated functions. In digital transformation, diverse backgrounds may broaden information access and agility while raising coordination costs through cognitive friction, producing complex moderating effects. Yet, existing studies focus predominantly on direct effects, leaving TMT heterogeneity's moderating role in the digital transformation–market value relationship systematically unexamined.

To address this lacuna, this study positions TMT heterogeneity as a contingent moderator within the digital transformation–market value framework. The dissonance between macro valuation trends and empirical conclusions, compounded by the absence of rigorous moderating analysis of TMT heterogeneity, constitutes a significant theoretical gap, as do the differential mechanisms across heterogeneity dimensions. Accordingly, this paper employs a quantitative longitudinal panel of Chinese A share listed firms from 2012 to 2022. Using fixed effects regression, we test the digital transformation–market value relationship and introduce TMT background heterogeneity as a moderator. This inquiry aims to unpack the operative pathways from an internal governance perspective, extend Upper Echelons Theory's explanatory reach into digital transformation, and provide theoretical and empirical grounding for corporate governance practice.

1.2. Research Purpose

This study examines how digital transformation influences firm market value, focusing on the moderating role of top management team (TMT) heterogeneity. It pursues three objectives: (1) empirically test the direct relationship between digital transformation

and market value; (2) systematically analyze the overall moderating effect of TMT heterogeneity within this nexus; and (3) decompose TMT heterogeneity into multiple dimensions, including age, educational attainment, functional background, and overseas experience, to reveal the boundary conditions and pathways through which distinct attributes shape value conversion, thereby explaining inter-firm variation in digital outcomes.

1.3. Research Significance

1.3.1. Theoretical Significance

Drawing on Upper Echelons Theory, this study integrates TMT heterogeneity into the digital transformation and market value framework, extending theoretical boundaries of value realization in digital contexts. While prior work emphasizes direct effects or singular moderators, this study adopts a multidimensional perspective to delineate how different heterogeneity dimensions exert differential moderating influences. This deepens understanding of the "digital paradox" and enriches Upper Echelons Theory's explanatory power in digital settings, offering novel evidence on how TMT composition shapes digital strategy efficacy.

1.3.2. Practical Significance

The findings guide firms in optimizing TMT structures to enhance digital transformation effectiveness. They inform the alignment of executive background configurations with strategic positioning and development stage, enabling firms to leverage heterogeneous teams' complementary strengths and improve value conversion efficiency. The results also provide empirical grounding for listed companies to refine governance mechanisms and executive selection, advancing high-quality corporate development.

1.4. Research Methods

1.4.1. Literature Review Method

This study systematically reviews domestic and international literature across digital transformation, firm market value, and TMT heterogeneity, following the logic of "theoretical evolution, then mechanisms, and finally research gaps." It synthesizes Upper Echelons Theory, digital value creation mechanisms, and the multidimensional effects of TMT heterogeneity, identifying consensus, contention, and the lacuna regarding TMT heterogeneity's moderating role, thereby establishing the foundation for framework construction and hypothesis development.

1.4.2. Empirical Research Method

A quantitative longitudinal panel design is employed using Chinese A-share listed firms from 2012 to 2022, constructing an unbalanced panel dataset. Multivariate regression tests the relationship between digital transformation and market value, as well as TMT heterogeneity's moderating effect. The panel design is justified by three considerations: (1) it captures temporal lags inherent in digital transformation outcomes;

(2) fixed-effects models control for time-invariant unobserved firm heterogeneity, mitigating endogeneity; and (3) larger sample sizes and within-variable variation facilitate rigorous testing of interaction terms and differential moderating effects across heterogeneity dimensions.

1.5. Innovations

1.5.1. Novel Research Perspective

While prior studies examine the relationship between digital transformation and firm value primarily through technology adoption, organizational change, or external environment lenses, this study shifts focus to internal governance by integrating Upper Echelons Theory and positioning TMT heterogeneity as a key moderating factor. This perspective illuminates the underlying drivers of divergent digital transformation outcomes and broadens the theoretical scope of the field.

1.5.2. Deepened Research Content

Existing work offers preliminary evidence on the link between TMT heterogeneity and digital transformation but seldom tests its moderating role within the broader framework of digital transformation and market value. This study addresses that gap by examining both the aggregate moderating effect and the differential impacts of distinct heterogeneity dimensions, including age, educational attainment, functional background, and overseas experience, thereby providing nuanced empirical insight into how specific executive attributes shape digital strategy effectiveness.

2. Theoretical Basis and Research Hypotheses

2.1. Digital Transformation and Corporate Market Value

Michael Porter's value chain theory provides a foundational framework for understanding the impact of digital transformation on firm value. This theory categorizes enterprise activities into primary and support activities, positing that their interconnections constitute a value creation system into which digital technologies can be embedded to substantially enhance efficiency and synergy across all stages. In the digital era, this logic has been further reinforced. By deeply integrating digital technologies with actual business processes, spanning R&D and design, manufacturing, operations management, and sales and service, firms can reconfigure product systems, production workflows, organizational structures, and business models (Xu Haoqing et al., 2024). Lai Hongbo (2020), through case analyses of Alibaba's Xunxi Digital Factory and Pinduoduo, demonstrates that digital transformation serves as a core enabler of retail supply chain innovation by digitizing production, distribution, and sales while integrating online and offline channels with logistics systems [9]. Liu Bo's (2022) empirical study of Chinese retail firms reveals that artificial intelligence and digital technologies significantly enhance firm value by advancing supply chain upgrading and precisely matching consumer demand [10]. Chen (2024) noted that digital transformation enables massive data use, facilitating precise market positioning and product design to boost competitiveness and market value [11]. Zhang Yuanji and

Han Cun (2023), examining A-share "Specialized, Refined, Distinctive, and Innovative" (SRDI) enterprises, find that digital transformation elevates firms' technological innovation capacity and innovation investment, thereby positively influencing market value [12]. Xu Haoqing et al. (2024), employing Sobel tests and bootstrapping methods on Chinese listed firms, demonstrate that digital transformation drives market value growth through improved productivity and innovation capability [13]. Li Yuejiao, utilizing the Qwen large language model to construct firm-level digital transformation indicators, empirically tests and confirms a significant positive effect of digital transformation on firm market value [14]. Synthesizing the foregoing theoretical and empirical evidence, this study contends that digital transformation, by leveraging digital technologies to optimize the efficiency and synergy of both primary and support activities, exerts a positive facilitating influence on firm market value. Accordingly, the following hypothesis is proposed:

H1: Digital transformation can enhance corporate market value.

2.2. TMT Heterogeneity, Digital Transformation, and Corporate Market Value

Upper Echelons Theory (Hambrick & Mason, 1984) posits that top management team (TMT) attributes such as age, gender, educational attainment, tenure, and functional background shape executives' cognitive orientations and interpretive schemas, thereby influencing strategic behaviors and organizational performance [15]. However, extant research remains divided on whether TMT heterogeneity yields positive or negative effects. Information processing theory suggests that heterogeneous teams possess broader information, knowledge, and capabilities, facilitating more comprehensive strategic decision making (Van Knippenberg et al., 2004) [16], a view supported by evidence that TMT heterogeneity positively affects firm growth (Su Taoyong et al., 2021) [17]. Conversely, social categorization theory emphasizes that demographic dissimilarities trigger in group favoritism and out group derogation, engendering intergroup biases that impair communication and coordination (Bhappu AD, 1997) [18]. Aligned with this perspective, Zhang Cheng (2023) finds that heterogeneity in age, tenure, and functional background impedes corporate digital transformation [19], Wu Chengcheng and Li Xuan (2024) report that age heterogeneity inhibits digital transformation in manufacturing firms [20], and Deng Xinming et al. (2021) document a negative effect of tenure heterogeneity on market performance [21]. These conflicting findings indicate that the consequences of TMT heterogeneity are highly contingent upon strategic context. Within the context of digital transformation, the adverse effects of heterogeneity are likely to be amplified. From a resource integration perspective, diverse professional backgrounds create knowledge silos that raise coordination costs and hinder the cross functional knowledge sharing essential to digital initiatives (Barney, 1991) [22]. From a decision making standpoint, executives from distinct functional origins tend to advocate subunit interests, and in high risk digital projects, such divergence intensifies goal conflict and risk averse resistance (Jensen & Meckling, 1976) [23]. Furthermore, cognitive fragmentation among heterogeneous members impedes consensus on the direction, pace, and magnitude of transformation efforts. Collectively, these mechanisms suggest that in the digital transformation setting, TMT heterogeneity attenuates the positive effect of digital transformation on firm market value by elevating coordination costs, provoking risk averse conflicts, and fostering strategic incoherence. Accordingly, the following hypothesis is proposed:

H2: TMT heterogeneity inhibits the promoting effect of digital transformation on corporate market value.

3. Research Methods and Experimental Results

3.1. Sample Selection and Data Sources

This study utilizes a sample of Chinese A-share listed firms spanning 2012–2022, with financial and operational data obtained from the CSMAR database. Observations with severe missing values are excluded, and all variables exhibit missing proportions below 5%, with mean comparisons between retained and deleted observations revealing no significant differences, thereby confirming that missingness is random and does not induce sample selection bias. To mitigate the influence of outliers, all continuous variables are winsorized at the 1st and 99th percentiles. Data processing is conducted using Stata 16.0, yielding a final sample of 20,439 firm-year observations. Complete variable definitions are provided in Table 1.

Table 1. Variable Definitions

Variable	Variable Name		Specific Calculation Method
Dependent Variable	Corporate Market Value	TobinQ	Market Value/Total Assets
Independent Variable	Digital Transformation Degree	Dzy	Sum of keyword frequencies related to digital transformation in annual reports
Moderating Variable	TMT Heterogeneity	TMT	Includes heterogeneity in age, tenure, overseas background, financial background, etc. Coefficient of Variation (CV) is used for continuous variables (e.g., age); Herfindal-Hirschman Index (HHI) is used for categorical variables (e.g., financial/overseas background), $H=1-\Sigma Pi^2$. Individual heterogeneity values are normalized and averaged to obtain the overall TMT heterogeneity index.
Control Variables	Firm Size	Size	Natural logarithm of total assets
	Retention Ratio	Syl	1-(Dividends per share before tax) /((Net profit current period/Paid-in capital end of period)
	Current Ratio	Blv	Current Assets/Current Liabilities
	Equity Turnover	Gd	Operating Income/ Ending Balance of Shareholders' Equity
	Return on Capital	Zbb	(Total Profit + Financial Expenses) / Total Assets
	Firm Age	Age	Ln (Years since establishment+1, as of 2023)
	Operating Leverage	Jy	(Net Profit +Income Tax+Financial Expenses+Depreciation & Amortization) / (Net Profit+Income Tax+Financial Expenses)
	Capital Accumulation Rate	Zjl	(Ending Total Equity - Beginning Total Equity) / Beginning Total Equity
	Duality	Dual	1 if Chairman and CEO are the same person; 0 otherwise
	Proportion of independent directors	DLD	Ratio of independent directors to board size
	R&D expenditure	RD	The ratio of a firm's annual R&D expenditure to its total assets
	Ownership nature	SOE	1 for state-owned enterprises, 0 otherwise

3.2. Variable Definition and Measurement

3.2.1. Dependent Variable

The dependent variable is Corporate Market Value (TobinQ). Compared to using accounting performance indicators, Tobin's Q better measures corporate performance and growth potential. Therefore, referring to Zhang and Zhu (2019) [24], this paper selects Tobin's Q as the dependent variable.

3.2.2. Core Independent Variable

The independent variable is Digital Transformation Degree (Dzy). There is no unified standard for measuring digital transformation. Referring to Wu et al. (2021), this paper selects keywords related to digital transformation in annual reports [25]. The degree is measured by counting 76 digital-related keywords across five dimensions: artificial intelligence, big data, cloud computing, blockchain, and digital technology application.

3.2.3. Moderating Variable

The moderating variable is TMT Heterogeneity. It includes age, tenure, overseas background, and financial background heterogeneity. Referring to Huang et al. (2011), the Coefficient of Variation (CV) is used for continuous variables (e.g., age), and the Herfindal-Hirschman Index (HHI) is used for categorical variables (e.g., financial/overseas background), where $H=1-\Sigma Pi^2$ (Pi is the proportion of the i-th category) [26]. Referring to Gao et al. (2024), individual heterogeneity values are normalized and averaged to obtain the overall TMT Heterogeneity index [27]. A higher value indicates higher TMT heterogeneity.

3.2.4. Control Variables

To avoid omitted variable bias, Firm Size (Size), Current Ratio (Blv), Total Asset Turnover (Zzl), Return on Capital (Zbb), Firm Age (Age), Financial Leverage (Gg), Capital Accumulation Rate (Zjl), Duality (Dual),Proportion of independent directors (DLD), R&D expenditure(RD) and Ownership nature(SOE) are included as control variables. Individual firm and year fixed effects are simultaneously controlled for in the model.

3.3. Research Model

To test the relationship between digital transformation and market value, the following baseline regression model is constructed:

$$TobinQ_{i,t}=\beta_0+\beta_1\ Dzy_{i,t}+\sum controls_{i,t}+\sum year+\sum industry+\varepsilon_{i,t}$$

Where TobinQ, Dzy, and Controls are the dependent, independent, and control variables, respectively. ε is the random error term. Year and individual dummy variables are included for robustness.

To test the moderating effect of TMT heterogeneity, the following model is constructed:

$$TobinQ_{i,t}=\beta_0+\beta_1 Dzy_{i,t}+\beta_2 TMT_{i,t}+\beta_3 Dzy_{i,t}\times TMT_{i,t}+\sum controls_{i,t}+\sum year+\sum industry+\varepsilon_{i,t}$$

The independent variable Dzy and the moderating variable TMT are first mean centered, after which the interaction term Dzy × TMT is introduced to test the moderating effect of top management team heterogeneity.

3.4. Scope of the Study

This study focuses exclusively on Chinese A-share listed firms, and the generalizability of its findings is accordingly bounded by this scope. Specifically, the results may not extend to unlisted firms, small and medium-sized enterprises, or firms operating within divergent institutional environments. Furthermore, digital transformation is measured via textual analysis of annual report disclosures, capturing the extent of digitalization-related discourse rather than direct capital investment. TobinQ is employed to measure firm market value rather than broader performance indicators.

4. Empirical Analysis

4.1. Descriptive Statistics

Table 2. Descriptive Statistics

Variable	Obs	Mean	Median	Std. Dev.	Min	Max
TobinQ	20439	2.00	1.61	1.20	0.85	7.687
Dzy	20439	12.87	2	28.50	0	181
TMT	20439	0	0	0.48	-1.74	1.78
TD	20439	0.78	0	13.28	-197.20	199.40
Syl	20439	0.69	0.75	0.31	-0.92	1
Gd	20439	1.22	0.90	1.10	0.14	6.81
Jy	20439	1.55	1.34	0.66	1.02	5.35
Zbb	20439	0.07	0.06	0.04	0	0.24
Zjl	20439	0.22	0.08	0.39	-0.09	2.37
Blv	20439	2.61	1.79	2.51	0.42	16.21
Size	20439	22.28	22.09	1.27	20.08	26.29
Age	20439	2.87	2.89	0.34	1.79	3.56
Dual	20439	0.30	0	0.46	0	1
DLD	20439	0.38	0.35	0.05	0.33	0.57
RD	20439	0.02	0.02	0.02	0	0.10
SOE	20439	0.31	0	0.46	0	1

Table 2 shows that the mean of digital transformation (Dzy) is 12.87, with a standard deviation of 28.50, indicating significant differences among listed companies. The mean of Tobin's Q is 2.00, with a standard deviation of 1.20, showing large variations in market value. The TMT heterogeneity data is normally distributed after processing. Other control variables are consistent with the actual situation of listed companies.

4.2. Multicollinearity Diagnostics

Table 3. Main Effect Regression

Variable	VIF	1/VIF
Size	1.40	0.715
SOE1	1.36	0.736
Zbb	1.35	0.738
Jy	1.34	0.745
Blv	1.31	0.761
RD11	1.22	0.818
Dzy	1.18	0.847
Gd	1.17	0.856
Dual	1.12	0.894
TMT	1.11	0.902
Age	1.08	0.930
Zjl	1.06	0.943
Syl	1.05	0.951
TD	1.03	0.972
DLD	1.02	0.979

To ensure that high inter-variable correlations do not bias subsequent regression and moderation analyses, a multicollinearity diagnostic was performed (see Table 3). The maximum variance inflation factor (VIF) is 1.40, with all variables, including the moderator and interaction terms, exhibiting low VIF values. These results confirm the absence of severe multicollinearity, establishing the data's suitability for the ensuing regression procedures.

4.3. Baseline Regression

Table 4. Main Effect Regression

Variable	(1) TobinQ	(2) TobinQ	(3) TobinQ	(4) TobinQ
Dzy	0.001*	0.001***	0.001***	0.001**
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	1.987***	8.522***	6.165***	6.614***
	(0.016)	(0.324)	(0.236)	(0.444)
Controls	No	Yes	Yes	Yes
Individual	No	Yes	No	Yes
year	No	No	Yes	Yes
Observations	20439	20439	20439	20439
R^2		0.083		0.313

Note: * p<0.1; ** p<0.05; *** p<0.01; R2 is unadjusted; Coefficients are standardized. The same below.

Correlation analysis and multicollinearity tests were conducted. All variables reject the correlation hypothesis, and VIF values are less than 10, indicating no multicollinearity. The Hausman test (p=0.000) rejects the null hypothesis, so a fixed-effects model is adopted. Model (1) has no controls, Model (2) has fixed individual effects, Model (3) has fixed time effects, and Model (4) has both.

According to Model (4), the coefficient of Dzy is 0.001 (p<0.05), significant at the 5% level, indicating that digital transformation promotes market value, supporting Hypothesis H1.

4.4. Moderating Effect

Table 5. Moderating Effect Regression

Variable	(1) TobinQ
Dzy	0.001**
	(0.000)
TMT	-0.040**
	(0.020)
TD	-0.001*
	(0.001)
Constant	6.612***
	(0.432)
Controls	Yes
Individual	Yes
year	Yes
Observations	20439
R^2	0.313

The coefficient of Dzy is 0.001. The interaction term (Dzy×TMT) has a coefficient of -0.001 ($p<0.01$), significantly negative at the 5% level. This indicates that TMT heterogeneity inhibits the positive effect of digital transformation on market value, supporting Hypothesis H2.

4.5. Robustness Test

A robustness test was conducted by replacing the core independent variable. Referring to Zhen et al. (2023) [28], the keyword frequency extraction method for Dzy was changed to count 139 digital-related keywords under categories such as technology classification, organizational empowerment, and digital application, defined as Dzy0.The coefficient of Dzy0 is 0.001 ($p<0.01$), consistent with the baseline regression. Table 6(2) shows that the interaction term (Dzy0×TMT) is -0.001 ($p<0.05$), consistent with the moderating effect results. Additionally, a robustness test was conducted by truncating the sample period from the original 2012–2022 range to 2012–2019. As shown in columns (3) and (4), the results for both the baseline regression and the moderating effect remain consistent with the original findings. Thus, H1 and H2 remain valid.

Table 6. Robustness Test

Variable	(1) TobinQ	(2) TobinQ	(3) TobinQ	(4) TobinQ
Dzy0	0.001**	0.001**	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
TMT		-0.016		-0.025
		(0.021)		(0.023)
TD1		-0.001*		-0.001**
		(0.000)		(0.001)
Individual	Yes	Yes		
Year	Yes	Yes		
Constant	6.971***	6.964***	9.299***	9.286***
	(0.432)	(0.432)	(0.570)	(0.571)
Controls	Yes	Yes	Yes	Yes
Observations	20904	20904	14453	14453
R^2	0.308	0.308	0.388	0.388

4.6. Endogeneity Test

Table 7.Endogeneity Test

Variable	(1) TobinQ
Dzy	0.000941*
	(2.52)
Controls	Yes
N	14805

A salient concern in the digital transformation–market value relationship is potential reverse causality, whereby higher market valuations may furnish greater financial resources and investor pressure, leading firms to invest more in digitalization rather than the reverse. To mitigate this concern, an endogeneity test was conducted by lagging the independent variable by one period. As reported in Table 7, the lagged independent and control variables significantly predict current period firm performance, and the results align closely with the primary findings, indicating that the relationship is not merely reciprocal. The positive association between digital transformation and market value remains statistically significant at the 5% level, demonstrating the model's satisfactory resilience to endogeneity concerns.

4.7. Heterogeneity Test

To further explore the role of different dimensions of TMT background heterogeneity, this paper divides it into age (TMT1), gender (TMT2), educational background (TMT3), professional background (TMT4), overseas background (TMT5), and financial background (TMT6).

Results in columns (3) and (6) (non-significant results are hidden) show that the coefficient of educational background heterogeneity (TMT3) is -0.136 ($p<0.05$), indicating an inhibitory effect on market value. The interaction term of financial background heterogeneity (TMT6) and Dzy (TTT6) is -0.003 ($p<0.1$), significant at the 10% level. This suggests that financial background heterogeneity significantly influences the promotion of digital transformation on market value. In summary, financial background heterogeneity is the main factor contributing to the inhibitory effect of overall TMT heterogeneity.

Table 8. Heterogeneity Test

Variable	(1) TobinQ	(2) TobinQ	(3) TobinQ	(4) TobinQ	(5) TobinQ	(6) TobinQ
Dzy	0.001**	0.001**	0.001**	0.001**	0.001**	0.001**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
TMT3			-0.136***			
			(0.044)			
TTT3			-0.002			
			(0.002)			
TMT6						0.017
						(0.060)
TTT6						-0.003*
						(0.002)
Constant	6.614***	6.608***	6.615***	6.705***	6.590***	6.623***
	(0.444)	(0.444)	(0.444)	(0.445)	(0.444)	(0.444)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Individual	Yes	Yes	Yes	Yes	Yes	Yes

year	Yes	Yes	Yes	Yes	Yes	Yes
Observations	20904	20904	20904	20904	20904	20904
R^2	0.307	0.307	0.308	0.308	0.307	0.308

5. Conclusion

5.1. Research Conclusions

Using 2012-2022 data of Chinese A-share listed companies, this paper explores the relationship among TMT heterogeneity, digital transformation, and market value. Key findings: (1) Digital transformation degree boosts market value; (2) TMT heterogeneity exerts a moderating effect, inhibiting digital transformation's positive impact on market value; (3) Educational background heterogeneity also suppresses market value; among TMT heterogeneity dimensions, financial background heterogeneity is the primary inhibitory factor.

5.2. Management Implications

Digitalization is critical for enterprises. They should align with digital trends, leverage big data, AI, and cloud computing per business needs to deepen digital transformation, optimize resource allocation, and adjust R&D investment dynamically to enhance market value.

Policy Implications: The government has issued pro-digital transformation policies. Future efforts should include more supportive policies, advancing digital-tech integration with the real economy, and improving supporting measures to drive high-quality development.

Corporate Governance Implications: Given TMT heterogeneity's inhibitory effect, enterprises should conduct rigorous executive background checks. Heterogeneity dimensions (professional, educational, financial) have varying impacts; financial background heterogeneity may cause risk-return conflicts, hindering digital transformation. Enterprises should optimize TMT structure by development stage to avoid excessive heterogeneity impeding digital value creation.

References

[1] Song J, Huang HJ, Jiang YS. Research on the impact of digital transformation on corporate ESG performance. Soft Sci. 2024. DOI: 10.13956/j.ss.1001-8409.2024.00.000

[2] Wang D, Ma ZQ, Xu LY. Digital transformation and corporate value creation: A dual perspective of product and service innovation. Sci Technol Prog Policy. 2024. DOI: 10.6049/kjjbydc.2023060626

[3] Zhang XX, Dong Y, Yi T. Digital transformation and sustained innovation value appropriation: A new perspective based on relational capital. J Shanxi Univ Financ Econ. 2024;46(11):113-126. DOI: 10.13781/j.cnki.1007-9556.2024.11.009

[4] Wang XH, Luan XY, Zhang SP. Corporate R&D investment, ESG performance and market value: The moderating effect of corporate digitalization level. Stud Sci Sci. 2023;41(05):896-904+915. DOI: 10.16192/j.cnki.1003-2053.2023.05.012

[5] Huang DY, Xie HB. Research on the impact and mechanism of corporate digital transformation on market value: Empirical evidence from Chinese listed companies. Asia Pac Econ Rev. 2022;(06):93-104. DOI: 10.13654/j.cnki.napee.2022.06.009

[6] Mason A, Carpenter MA, Golden BR. Upper echelons research revisited: Antecedents, elements, and consequences of top management team composition. J Manag. 2004;30(6):749-778. DOI: 10.1177/0149206304263071

[7] Han QX, Yang C, Gu ZP. Threshold effect of top management team heterogeneity on corporate innovation efficiency: An empirical study based on listed companies in strategic emerging industries. Issues Chin Econ. 2017;(02):42-53. DOI: 10.14007/j.cnki.cjpl.2017.02.005
[8] Tang X, Gao X, Zhao TQ, et al. Top management team heterogeneity and corporate digital transformation. China Soft Sci. 2022;(10):83-98. DOI: 10.3969/j.issn.1005-0566.2022.10.009
[9] Lai HB. The innovation mechanism of digital technology empowerment and "new retail": A case study of Alibaba Rhinoceros and Pinduoduo. China Circ Econ. 2020;34(12):11-19. DOI: 10.14089/j.cnki.cn11-3664/f.2020.12.002
[10] Liu B. Research on the impact of digital transformation of retail enterprises on corporate value improvement. J Bus Econ. 2022;(17):130-133. DOI: 10.3969/j.issn.1002-5863.2022.17.032
[11] Chen L. Research on the changing trend of corporate market value under the background of digital transformation. Market Modernization. 2024;(23):155-157. DOI: 10.3969/j.issn.1006-3102.2024.23.051
[12] Zhang YJ, Han C. Corporate digital transformation, technological innovation and market value: Empirical evidence from "specialized, sophisticated, distinctive and innovative" listed enterprises. Stat Decis. 2023;39(14):163-167. DOI: 10.13546/j.cnki.tjyjc.2023.14.026
[13] Xu HQ, Ma YF, Yang XW. Research on the impact of digital transformation on corporate market value. Study Explor. 2024;(03):95-103. DOI: 10.16513/j.cnki.1004-7794.2024.03.010
[14] LI Y J, SHI G F. The Impact of Digital Transformation on Corporate Market Value: An Analysis Based on Large Language Models. Taxation and Economy. 2026;9(2):193-206.
[15] Hambrick DC, Mason PA. Upper echelons: The organization as a reflection of its top managers. Acad Manag Rev. 1984;9(2):193-206. DOI: 10.5465/amr.1984.4282599
[15] Wu Y, Ding L, Song X, et al. Top management team heterogeneity and the performance of cross-border M&A. Sage Open. 2023;13(2):21582440231175127. DOI: 10.1177/21582440231175127
[16] Van Knippenberg D, De Dreu CK, Homan AC. Work group diversity and group performance: An integrative model and research agenda. J Appl Psychol. 2004;89(6):1008-1022. DOI: 10.1037/0021-9010.89.6.1008
[17] Su TY, Mao YF, Shan ZW. Top management team heterogeneity, ambidextrous innovation and corporate growth: The moderating effect of industry competition and slack resources. Sci Manag Res. 2021;39(06):75-81. DOI: 10.19445/j.cnki.15-1103/g3.2021.06.013
[18] Bhappu AD, Griffith TL, Northcraft GB. Media effects and communication bias in diverse groups. Organ Behav Hum Decis Process. 1997;70(3):199-205. DOI: 10.1006/obhd.1997.2695
[19] Zhang C. The impact of top management team heterogeneity on corporate digital transformation: The moderating role of ownership structure [dissertation]. Shanghai: Shanghai International Studies University; 2023. DOI: 10.27312/d.cnki.gshsu.2023.001234
[20] Wu CC, Li X. Financial slack, top management team heterogeneity and corporate digital transformation: Empirical evidence from listed manufacturing companies in the Yangtze River Delta. J Shijiazhuang Tiedao Univ (Soc Sci Ed). 2024;18(01):34-42. DOI: 10.13319/j.cnki.sjztddxxbskb.2024.01.06
[21] Deng XM, Luo H, Long XY, et al. Top management team heterogeneity, competitive strategy portfolio and market performance: An empirical test from China's household appliance industry. Nankai Bus Rev. 2021;24(04):103-117. DOI: 10.16538/j.cnki.nkij.2021.04.009
[22] BARNEY J. Firm resources and sustained competitive advantage. Journal of Management. 1991;17(1):99-120. DOI: 10.1177/014920639101700108
[23] JENSEN M C，MECKLING W H. Theory of the firm：Managerial behavior，agency costs and ownership structure. Journal of Financial Economics. 1976;3(4):305-360. DOI: 10.1016/0304-405X(76)90026-X
[24] Zhang XM, Zhu JG. Does operating asset structure affect the market value of high goodwill enterprises? An empirical study based on A-share listed companies. Nankai Bus Rev. 2019;22(02):114-127. DOI: 10.16538/j.cnki.nkij.2019.02.010
[25] Wu F, Hu HZ, Lin HY, et al. Corporate digital transformation and capital market performance: Empirical evidence from stock liquidity. J Manag World. 2021;37(07):130-144+10. DOI: 10.19744/j.cnki.11-1235/f.2021.0091
[26] Huang Y, Yang ND, Zhang CL. Research on the impact of top management team heterogeneity on corporate performance: With ownership concentration as a moderating variable. Manag Rev. 2011;23(11):120-125+168. DOI: 10.14120/j.cnki.cn11-5057/f.2011.11.015
[27] Gao Y, Zhang Y, Liu CJ. Can digital transformation improve corporate environmental, social and governance performance? The moderating role of top management team heterogeneity. Sci Technol Prog Policy. 2024;41(11):55-66. DOI: 10.6049/kjjbydc.2023090587
[28] Zhen HX, Wang XF, Fang HX. Administrative protection of intellectual property rights and corporate digital transformation. Econ Res J. 2023;58(11):62-79. DOI: 10.12088/PKU.jjkx.2023.11.04

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261074

A Study on Generation Z's Information Reception Behaviour Regarding Pre-prepared Meals in an Algorithm-Driven Recommendation Context

Zengjingwen Chen[a], Yufang Peng[a,b,1], Yuchen Zhang[a], Bei Liu[a] and Qi Liu[a]
[a.] *School of Management Engineering, Nanjing Institute of Technology, Nanjing 211167, China*
[b] *School of Information Management, Nanjing University, Nanjing 210023, China*

Abstract. In the short-video environment deeply embedded with algorithmic recommendations, this study explores Generation Z's information acceptance mechanisms regarding pre-prepared meals. Based on an analytical framework of 'algorithmic environment→risk content→cognitive processing→acceptance behaviour', we conduct regression analysis on 469 questionnaires and sentiment analysis on 3,684 Douyin comments. The findings reveal that algorithms only constitute the backdrop for information exposure; content quality, source credibility and information transparency are the key influencing factors. Comment analysis further shows that users' focus on 'value fairness' and 'the right to know' aligns with the policy orientation for mandatory labelling of pre-prepared meals. Public information acceptance results from the synergistic interaction between information environment, risk perception and trust-building, rather than being determined by a single technological factor.

Keywords. Algorithmic recommendation context, information behaviour, short video content, pre-prepared meals, sentiment analysis

1. Introduction

Algorithmic recommendation mechanisms have become deeply embedded in the short-video environment, influencing users' information exposure and consumption decisions. As an emerging food category tailored to fast-paced lifestyles, the market expansion of pre-prepared meals relies heavily on information distribution via platforms such as Douyin. As digital natives, Generation Z's consumption decisions are increasingly shaped by short-video content. However, existing research lacks

[1] Corresponding Author, Yufang Peng, Email: laisitianshi@163.com
This research was supported by the Young Foundation of Ministry of Education Project of Humanities and Social Sciences in China (Grand Nos: 22YJC870012), the School Research Foundation Project of Nanjing Institute of technology in China(Grand Nos: YKJ202231), the 2024 Labor Education Practice Project of Nanjing Institute of Technology in China (Grand Nos: LDSJ202440), the 2024 Central Government Special Fund for Supporting the Reform and Development of Local Universities (Humanities and Social Sciences Research Program) in China, and the enterprise-commissioned service project in China (Grand Nos: 25221).

systematic exploration of the pathways linking algorithmic mechanisms, content formats, and information acceptance among this demographic.

This study adopts a tripartite perspective, defining algorithms as the information environment, pre-prepared meals as the subject of risk information, and Generation Z as the key audience. Employing a combined methodology of questionnaire surveys and sentiment analysis, we examine how short-video content characteristics, credibility and presentation correlate with Generation Z's risk perception, trust-building and information acceptance intentions.

This study has clear policy implications. In 2024, six national ministries defined pre-prepared meals for the first time, emphasising the requirement of 'no preservatives' and promoting explicit labelling in catering operations. In 2026, the State Council's Food Safety Office launched a public consultation on voluntary disclosure of food preparation methods. These policies are closely linked to food safety transparency and consumers' right to know—core issues in digital risk communication. Global data from TelmarHelixa (2026) shows that 68% of Generation Z's food consumption information comes from algorithm-recommended short videos, supporting the rationale for this study.

This paper addresses three research questions:

(1) Do the dimensions of algorithmic perception directly influence Generation Z's information acceptance?

(2) What are the direct effects of short-video content characteristics and source credibility on information acceptance?

(3) What are the indirect effects and core transmission pathways via the chain of 'algorithm environment→content/source/attitude→acceptance'?

2. Relevant Research and Theoretical Foundations

2.1 Research on Short Video Marketing

A keyword clustering analysis of Chinese and English literature using CiteSpace revealed that "algorithm-based recommendations" have not yet formed an independent cluster, nor do they exhibit a clear strong association with keywords such as "food-related short videos" and "Generation Z" (see Figures 1 and 2).This suggests that existing research primarily views algorithms as technical tools for enhancing content distribution efficiency, whilst attention to the dynamic interactive mechanisms between the algorithmic environment, content characteristics and user cognition remains insufficient. Therefore, this study adopts the analytical pathway "algorithmic environment → content/source/attitude → acceptance" to attempt to address the shortcomings in existing research regarding the integration of algorithmic contexts and audience acceptance mechanisms.



Figure 1. Chinese Literature Keywords Cluster Map



Figure 2. Cluster map of keywords in foreign literature

2.2 Literature Review

Domestic research on pre-prepared meal consumption identifies convenience and perceived value as core drivers, whilst safety risks, additives and lack of transparency inhibit consumption willingness. Overseas research focuses on multidimensional perceived value and risk-value trade-offs. However, both streams treat algorithms as a static backdrop, failing to integrate the short-video algorithmic context or explain Generation Z's 'high awareness, low acceptance' phenomenon. This study addresses this gap by investigating the algorithm–content–user interaction mechanism.

2.3 Theoretical Basis

This study is underpinned by the SOR model and Perceived Value Theory. In the SOR model, algorithmic recommendations and short-video content constitute the stimulus (S); content preferences, source trust and risk perception constitute the organism (O); and information acceptance behaviour constitutes the response (R). Perceived Value Theory explains users' trade-off between benefits (convenience, emotional satisfaction) and drawbacks (safety, price, lack of transparency).

Grounded in the algorithmic recommendation context, this study extends the SOR model [1]: algorithm perception is decomposed into three quantifiable dimensions (matching, responsiveness, echo chamber) as explicit environmental stimuli; information transparency is introduced as an independent risk stimulus; and the chained mediation pathway of 'algorithm perception → content/trust/attitude → information acceptance' is validated. Consequently, this paper constructs an analytical framework comprising 'algorithmic environment → risk content → cognitive processing → acceptance behaviour'.

Based on the above, the following research hypotheses are proposed [2]:

H1: The various dimensions of algorithm perception are positively correlated with Generation Z's information acceptance behaviour regarding ready-made meals

H1a: Perceived algorithmic matching is positively correlated with Generation Z's information acceptance behaviour regarding ready-made meals; H1b: Perceived algorithmic responsiveness is positively correlated with Generation Z's information acceptance behaviour regarding ready-made meals; H1c: Perceived information echo chambers are positively correlated with Generation Z's information acceptance behaviour regarding ready-made meals.

H2: The various dimensions of perceived short video content characteristics are positively correlated with Generation Z's acceptance behaviour towards ready-made meal information

H2a: Informativeness is positively correlated with Generation Z's acceptance behaviour towards ready-made meal information; H2b: Utilitarianism is positively correlated with Generation Z's acceptance behaviour towards ready-made meal information; H2c: Experiential/Authenticity is positively correlated with Generation Z's acceptance behaviour towards ready-made meal information; H2d: Emotional appeal is positively correlated with Generation Z's acceptance behaviour towards ready-made meal information; H2e: Perceived information transparency is positively correlated with Generation Z's acceptance behaviour towards ready-made meal information.

H3: Organism-level cognition and affective evaluation are significantly correlated with Generation Z's acceptance behaviour towards ready-made meal information

H3a: Source credibility is positively correlated with Generation Z's acceptance behaviour towards ready-made meal information; H3b: Perceived functionality/value is positively correlated with Generation Z's acceptance behaviour towards ready-made meal information; H3c: Perceived risk is negatively correlated with Generation Z's acceptance behaviour towards ready-made meal information; H3d: Overall attitude is positively correlated with Generation Z's acceptance behaviour towards ready-made meal information.

H4: Variables at the organismal level may mediate the relationship between algorithm perception and Generation Z's acceptance behaviour towards ready-made meal information

Content preferences, source credibility and overall attitude act as mediators between the various dimensions of algorithm perception and Generation Z's acceptance behaviour towards ready-meal information; that is, algorithm perception indirectly influences their acceptance behaviour by affecting users' content preferences, source trust and overall attitude.

H5: Generation Z's media and consumption characteristics may be positively correlated with the influence pathways of ready-made meal information acceptance

Generation Z's media and consumption characteristics, such as a propensity for novelty-seeking and social sharing, reinforce the positive correlation between the perception of content features, cognitive evaluation and emotional evaluation, and their acceptance behaviour regarding ready-made meal information.

3. Data Sources and Research Methods

3.1 Data Sources and Data Preprocessing

This study employs a combined approach of questionnaire surveys and text mining. The questionnaire was distributed online via the WJX platform; 469 valid responses were obtained from Generation Z (born 1995-2009) who had encountered pre-prepared meal content on short-video platforms. Text data was sourced from Douyin, with 3,684 valid user comments collected after cleaning. (Data collection adhered to online ethical standards; user information was anonymised.)

3.2 Research Methods

This study adopts an 'algorithm-content-user' triadic approach. Quantitative analysis used SPSS for descriptive statistics, reliability/validity testing, correlation analysis and multiple regression. Text analysis employed dictionary-based sentiment analysis to extract semantic keywords, determine sentiment tendencies and analyse regional differences. The two data sets cross-validate each other; the research pathway is illustrated in Figure 3.

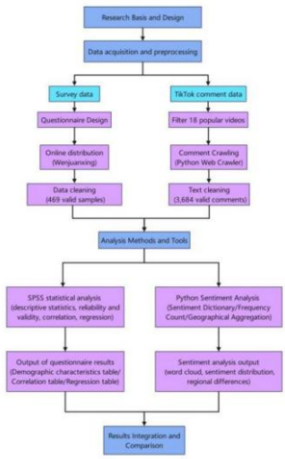


Figure 3. Research Technology Roadmap

4. Questionnaire Analysis of Generation Z's Acceptance of Pre-prepared Meals Based on SPSS

To accurately capture Generation Z's logic regarding the acceptance of ready-made meals within an algorithm-recommended environment, this section focuses on

quantitative research. Data was collected via a questionnaire survey system and analysed using SPSS, providing empirical support for revealing the intrinsic link between short-video marketing and acceptance.

4.1 Questionnaire Design

This study utilised a self-designed questionnaire for data collection. The questionnaire was constructed based on the SOR model and perceived value theory, aiming to systematically examine the influence pathways of algorithmic recommendation mechanisms, information content characteristics, and user psychological cognition on the acceptance of ready-made meal information.

To control common method bias, procedural measures were implemented during questionnaire design: neutral wording, randomised item ordering, anonymous completion, reverse-scored items, and response duration screening.

4.2 Data Collection and Sample Characteristics

The survey was distributed online via the Wenjuanxing platform, yielding a total of 513 responses. After screening and excluding invalid data, 469 valid responses were obtained, representing a response rate of 91.4%. All respondents were members of Generation Z, born between 1995 and 2009, and had all encountered content related to ready-to-cook meals on platforms such as Douyin, Kuaishou, Xiaohongshu, Bilibili or WeChat Video Accounts. The data was entered into SPSS for cleaning and analysis. Variable naming followed standard conventions, and the proportion of missing values was below 1%; these have been deleted.

4.3 Analysis of Questionnaire Results

4.3.1 Descriptive Statistics

(1) Demographic Characteristics of the Sample

Among the valid sample, females accounted for 53.7% and males for 46.3%, indicating a relatively balanced gender distribution. In terms of status, employed individuals (46.5%) and students (39.4%) constituted the two main groups, together accounting for 85.9% of the sample, which closely aligns with the profile of Generation Z. In terms of financial circumstances, 70.3% of respondents had a monthly disposable income of between 1,001 and 4,000 yuan, consistent with the income characteristics of students and those newly entering the workforce. It is worth noting that 33.9% of respondents rarely cook (0–1 times per week), whilst nearly 60% cook no more than three times a week. This finding aligns with the trend of reduced cooking activity and a pronounced desire to save time among younger demographics, and provides a contextual reference for understanding the market penetration of ready-to-cook meals among Generation Z.

(2) Short Video Usage Habits and Attitudes Towards Ready-to-Cook Meals

Regarding short video usage, Generation Z exhibits multi-platform and high-intensity characteristics. Douyin has the highest usage rate (75.7%), with 59.5% of respondents viewing short videos for more than two hours daily. In terms of exposure to ready-made meal content, 45.8% are in a state of high-frequency exposure, and 82.9% have had purchasing experience. Regarding algorithm perception and perceptions of

ready-made meals, the highest score was for perceived algorithmic responsiveness (M=3.69), with risk perception (M=3.66) slightly higher than functional perception (M=3.54), whilst the lowest score was for information acceptance intention (M=3.33), indicating a gap between cognition and behavioural conversion.

4.3.2 Reliability and Validity Testing

To comprehensively test the internal consistency reliability, convergent validity and discriminant validity of the measurement instrument, and to address the potential issue of an excessively high overall reliability coefficient, this study sequentially conducted reliability analysis, exploratory factor analysis (EFA) and confirmatory factor analysis (CFA), whilst employing Harman's single-factor test to assess common method effects.

(1) Internal Consistency Reliability

The total scale Cronbach's α was 0.969, with latent variable α ranging from 0.782 to 0.915, and composite reliability (CR) from 0.789 to 0.912—all exceeding the 0.70 threshold (Table 1). The high total α is attributable to reasonable theoretical correlations between latent variables.

Table 1. Results of reliability analysis for each latent variable

latent variable construct	Number of items	Corresponding questionnaire items	Cronbach's α coefficient	Composite reliability (CR)	Reliability assessment
Algorithm Matching Sense (AM)	2	Q9,Q10	0.803	0.876	good
Algorithmic Response and Interaction Sense (AR)	2	Q11,Q12	0.821	0.888	good
Information Cocoon Perception (IC)	2	Q13,Q14	0.780	0.863	good
Content type preference (CP)	4	Q15-Q18	0.897	0.894	good
Source credibility (SC)	2	Q19,Q20	0.752	0.854	good
Functional cognition (FC)	3	Q21-Q23	0.904	0.907	good
Risk Perception (RP)	3	Q24-Q26	0.917	0.914	good
Overall Attitude (AT)	2	Q27,Q28	0.805	0.899	good
Consumer Characteristics of Generation Z (ZT)	4	Q32-Q35	0.888	0.880	good
Information Acceptance Intention (IA)	3	Q29-Q31	0.906	0.903	good
Summary Table	29	Q7-Q35	0.969	-	good

(2) Construct Validity and Structural Validity

Exploratory factor analysis (EFA) showed KMO=0.979, Bartlett's test significant ($p<0.001$), with 11 common factors extracted accounting for 78.65% of variance. Confirmatory factor analysis (CFA) indicated good model fit ($\chi^2/df=2.356$,CFI=0.962,TLI=0.957,RMSEA=0.054,SRMR=0.048). Average variance extracted (AVE) values ranged from 0.526 to 0.812, all exceeding 0.50, confirming good convergent validity.

Table 2. Validation Factor Analysis Model Fit Index

fitness index	χ^2/degrees of freedom	CFI	TLI	RMSEA	SRMR
Modeling results	2.356	0.962	0.957	0.054	0.048
ideal criterion of judgment	≤3.0	≥0.90	≥0.90	≤0.08	≤0.08

(3) Discrimination validity

Heterotrait-Monotrait (HTMT) ratios ranged from 0.425 to 0.823 (Table 3), all below the 0.85 threshold, confirming good discriminant validity.

Table 3. Discriminant validity coefficient matrix of core latent variable HTMT

Latent variable	Algorithmic Matchability	Content Type Preference	Source Credibility	Functional Cognition	Risk Perception	Overall Attitude	Information Receiving Intention
Algorithmic matchability	-	0.756	0.689	0.712	0.658	0.792	0.805
Conten type preference	0.756	-	0.723	0.789	0.698	0.812	0.823
source credibility	0.689	0.723	-	0.654	0.612	0.765	0.765
Functional cognition	0.712	0.789	0.654	-	0.723	0.795	0.788
Risk perception	0.658	0.698	0.612	0.723	-	0.689	0.675
Overall attitude	0.792	0.812	0.765	0.795	0.689	-	0.818
Information receiving intention	0.805	0.823	0.765	0.788	0.675	0.818	-

(4) Common method effect test

Harman's single-factor test and subsequent analyses confirmed that common method bias is within an acceptable range.

4.3.3 Correlation Analysis

(1)Acceptance and Perceived Algorithm Performance, Content Preferences

Correlation analysis revealed that the intention to accept the information was significantly and positively correlated with both the perceived algorithm match $r = 0.616$ and the perceived algorithm responsiveness $r = 0.566$ ($p<0.01$).The correlation coefficients between the four types of content preference and acceptance intention ranged from 0.596 to 0.657, with utilitarian content showing the highest correlation $r = 0.657$, followed by informational $r = 0.625$ and emotional $r = 0.623$ content. This indicates that both algorithm perception and content preference are significantly and positively associated with Generation Z's willingness to accept ready-made meals [3].

(2)Acceptance and Risk Perception

The positive correlation between risk perception and acceptance intention ($r=0.589$, $p<0.01$) contradicts traditional consumer theory. This reflects Generation Z's decision-making logic in algorithmic environments: risk perception drives active information seeking, and algorithmic recommendations reinforce a 'risk-content-behaviour' chain, resulting in a 'cautious trial' pattern rather than outright rejection. This challenges the linear assumption that risk inevitably inhibits consumption [4].

4.3.4 Hierarchical Regression Analysis

Hierarchical regression was conducted with three models (Table 4). Model 1 included only algorithm perception dimensions; Model 2 added content and cognitive variables; Model 3 further added attitude and Generation Z traits.

Table 4. Results of Stratified Regression Analysis

Variable	Model 1 (β)	Model 2 (β)	Model 3 (β)	VIF
Algorithm Variable				
Algorithmic Matching	**0.394*****	**0.109***	**-0.016**	**3.72**
Algorithmic Response and Interaction	0.087	-0.099	-0.033	3.77
Information Cocoon Perception	0.232***	0.006	-0.002	2.35
Content and Cognitive Variables				
Content Type Preference		0.370***	0.187***	3.69
Source Trust		0.185***	0.151***	2.99
Functional Cognition		0.356***	0.156**	3.55
Risk Perception		-0.048	-0.087*	3.17
Attitude and Traits				
Overall Attitude			0.330***	2.90
The Consumption Characteristics of Generation Z			0.254***	3.70
model statistics				
R^2	0.419	0.655	**0.728**	
ΔR^2	**0.419**	0.236	0.073	
F	111.776***	78.674***	61.651***	

Note: p<0.001, p<0.01, p<0.05

The results of the regression analysis are shown in Table 4. The R^2 for Model 1 is 0.419 $F = 111.776, p < 0.001$, indicating that algorithm perception alone explains 41.9% of the variance in acceptance.Specifically, both the perception of algorithmic matching $\beta = 0.394, p < 0.001$ and the perception of information filter bubbles $\beta = 0.232, p < 0.001$ significantly and positively predicted acceptance, whilst the perception of algorithmic responsiveness and interactivity did not reach statistical significance.These results preliminarily indicate that users' perceptions of the alignment between algorithm-recommended content and their own interests, as well as their awareness of information narrowing, are both positively correlated with willingness to accept [5] .

After incorporating content and cognitive variables into Model 2, R^2=0.655, ΔR^2=0.236 (p<0.001), with the two variables collectively contributing 23.6% of the explanatory power. Content type preference $\beta = 0.370$, source trust $\beta = 0.185$ and functional cognition $\beta = 0.356$ all significantly and positively predicted acceptance (all $p < 0.001$);the coefficient for perceived algorithmic matching decreased from 0.394 to 0.109 $p = 0.035$, whilst the effects of the remaining algorithm-related variables were not significant, indicating that the relationship between perceived algorithmic matching and acceptance is mediated by content preferences and cognitive judgements.

After incorporating attitude and Generation Z traits into Model 3, R^2=0.728, ΔR^2=0.073 ($p < 0.001$). Both attitude $\beta = 0.330$ and Generation Z traits $\beta = 0.254$ significantly and positively predicted acceptance (both p<0.001);content preferences, source trust and functional cognition remained significant but with weakened effects, whilst the sense of algorithmic fit was entirely insignificant (β=−0.016,p=0.738), further confirming that the influence of the algorithm is absorbed by the mediation of proximal psychological variables.

Meanwhile, changes in the coefficient for risk perception are also noteworthy. In Model 2, risk perception did not show a significant effect (β=−0.048, p=0.315), whereas in Model 3, after incorporating attitude, its negative effect became significant (β=−0.087, p=0.044).This phenomenon of 'sign reversal' reveals the underlying mechanism linking attitude, risk perception and acceptance: whilst risk perception exerts an inhibitory effect on acceptance, this effect is masked by users'

positive attitudes—some users, despite harbouring safety concerns, recognise the product's convenience or hold a favourable view of the recommended content, with these positive factors diluting the negative impact of risk. Upon controlling for the attitude variable, the true negative effect of risk perception is revealed. The results indicate that Generation Z's consumer decision-making follows a dynamic risk-benefit trade-off logic, rather than a simple linear relationship of 'concern equating to rejection'.

Finally, focusing on the results of the multicollinearity diagnosis, the variance inflation factors for all variables were below 4, indicating that the model does not suffer from severe multicollinearity issues. The lack of significance of the algorithm variable in Model 3 is not due to multicollinearity, but rather because its explanatory power is absorbed by more proximal psychological variables such as content preference, source trust and attitude. This provides empirical support for the 'algorithm-content-user' tripartite interaction framework: algorithmic recommendations primarily constitute the environmental context of information exposure, whilst what truly drives information acceptance is content quality, source credibility and the user's cognition and attitudes formed on the basis of this information.

4.3.5 Testing for Mediating Effects

Bootstrap mediation analysis (Table 5) showed that content preference, source trust and attitude fully mediated the relationship between perceived algorithmic matching and acceptance (total indirect effect=0.5692, 95% CI [0.4949, 0.6454]). The indirect effects ranked by strength: attitude (0.2789)>content preference (0.1624)>source trust (0.1280). This confirms that algorithms influence acceptance indirectly through these proximal psychological variables.

Table 5. Results of Bootstrap Test for Mediation Effect

Way	Indirect Effect size	Boot SE	Boot LLCI	Boot ULCI
Total Indirect Effect	**0.5692**	0.0387	0.4949	0.6454
Algorithmic Matching→Content Preference→Acceptance	0.1624	0.0346	0.0949	0.2306
Algorithm Matching→Trust→Acceptance	0.1280	0.0304	0.0703	0.1905
Algorithm matching→Attitude→Acceptance	0.2789	0.0356	0.2050	0.3445
Direct Effect	0.0352	0.0362	-0.0359	0.1064
Gross Effect	**0.6044**	0.0358	0.5341	0.6748

Note: Bootstrap repeated sampling was performed 5000 times, with a 95% confidence interval using the bias correction percentile method.

5 Mining of Public Opinion Emotions and Focus Points of Pre-made Food on Douyin Platform Based on Sentiment Analysis

To compensate for the limitations of questionnaire data in revealing public discourse, this section analyses Douyin comments as an external supplement and contextual validation. Note that public comments do not allow identification of commenters' generation, so results reflect overall platform sentiment.

5.1 Data Sources

Python was used to scrape 18 popular pre-prepared meal videos (likes ≥ 10,000, comments > 3,000) and 4,298 user comments. After cleaning, 3,684 valid comments were retained.

5.2 Data Pre-processing

The raw comments underwent content review, and the following three categories of invalid data were excluded: (1) invalid comments with empty content or containing only emojis and special characters; (2) fragmented content with incomplete semantics or no clear viewpoint; (3) comments unrelated to the topic of ready-made meals. Following cleaning and filtering, 3,684 valid comments were retained as the corpus for subsequent sentiment analysis.

5.3 Sentiment Analysis Methods and Process

To ensure the reproducibility and verifiability of the sentiment analysis, this study employs a dictionary-based weighted sentiment scoring model, combining manual coding with word frequency statistics to complete the analysis. The specific process is as follows.

5.3.1 Core Methodology of Sentiment Analysis

A dictionary-based weighted sentiment model was employed [6], following a 'core dictionary + domain expansion' strategy [7]. The core lexicon integrated Hownet, Tsinghua University's Li Jun lexicon, and BosonNLP, supplemented with domain-specific terms via TF-IDF on 500 manually annotated comments. A weighting mechanism for degree adverbs and negation words was applied. Validation against a standard Chinese BERT model on 300 random samples showed 76.3% agreement (Kappa=0.65). Comments with scores>0.6 were classified as positive, <0.4 as negative, and otherwise neutral.

5.3.2 Dimensions for Topic and Sentiment Mining

Three mining dimensions were established:(1) semantic keyword extraction via word cloud; (2) overall sentiment classification (positive/neutral/negative); (3) regional sentiment analysis based on IP location data.

5.4 Sentiment Analysis Results

5.4.1 Core Topics and Keyword Characteristics

Word frequency analysis (Figure 4) revealed prominent terms: 'instant noodles', 'price', 'freshly cooked', 'fresh', 'cheap', 'cost', 'disclosure', 'fraud'. The high frequency of 'instant noodles' reflects users comparing it favourably to pre-prepared meals, which are criticised for being 'sold at a premium under the guise of being freshly cooked'. The

co-occurrence of economic and quality terms indicates the discussion centres on value fairness and the lack of transparency.

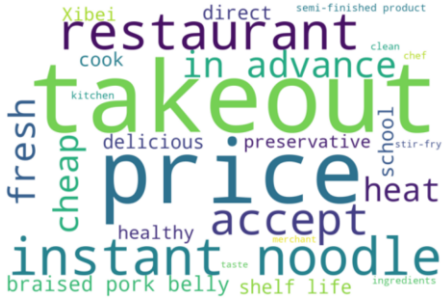


Figure 4. Word Cloud of Douyin Pre-made Meal Video Comments

5.4.2 Overall Sentiment Trends

As shown in Figure 5, negative comments accounted for 63.2% (2,328), positive for 25.1% (925), and neutral for 11.7% (431). Negative sentiment concentrated on three pain points: pricing concerns, lack of transparency, and food safety worries. Positive reviews appeared where prices were reasonable or products were clearly labelled as pre-prepared, suggesting users reject paying a 'freshly cooked premium' rather than rejecting pre-prepared meals themselves.

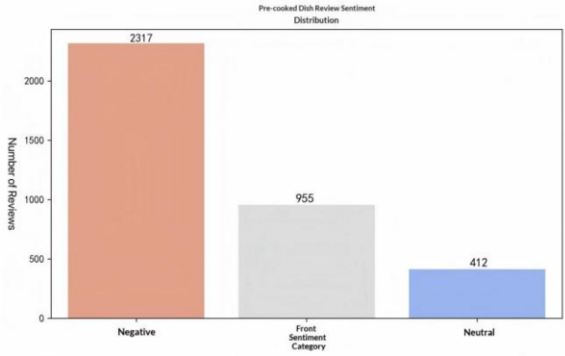


Figure 5. Emotional distribution map of Douyin pre-made meal video comments

5.4.3 Regional Sentiment Differences

Regional analysis (Figure 6) showed strongest negative sentiment in Northeast, Northwest and Southwest China, focusing on food safety and additives. East and Central China focused more on price and integrity issues. Guangdong and Henan, with strong catering industries, showed polarised views—from fierce criticism to rational acceptance. These regional differences indicate that risk perception and transparency demands vary across groups and regions.

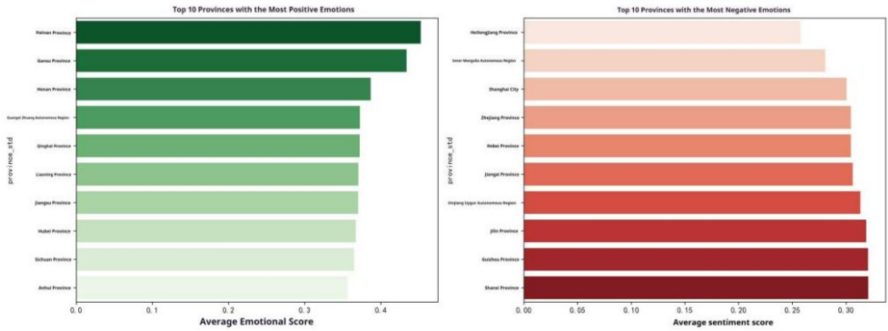


Figure 6. Analysis of regional differences in TikTok pre-made meal video comments

6 Conclusion and Prospects

6.1 Research Conclusions

(1) Generation Z's information acceptance follows a progressive 'cognition-attitude-intention' pattern, with attitude as the core driver. Descriptive statistics show 'high awareness, high concern, low acceptance'; hierarchical regression indicates overall attitude and Generation Z's consumption characteristics are the primary factors influencing acceptance intention.

(2) Algorithms serve as the backdrop for information exposure; content quality and source credibility are key to acceptance. Algorithmic relevance correlates positively with acceptance intention but is fully mediated by content preferences, source trust and attitude. Algorithms facilitate 'content discovery', but acceptance depends on content quality and risk-benefit trade-offs.

(3) Platform opinion exhibits a 'value fairness' orientation, with users focusing on price reasonableness and information transparency. Negative sentiment on Douyin (63.2%) centres on 'price fraud' and 'lack of transparency', not on rejecting pre-prepared meals. This aligns with national policy, suggesting user acceptance is higher when information is transparent and prices are reasonable.

The study validates the SOR model in algorithmic contexts: algorithmic perception indirectly influences acceptance through organism-level variables; information transparency shows a significant positive effect; the chained mediation pathway provides empirical evidence for risk information acceptance in digital environments.

6.2 Implications for Governance

The findings of this study offer the following implications for information governance on algorithmic platforms and food safety risk communication:

(1) Institutionalise information labelling on platforms to safeguard consumers' right to know.

(2) Establish risk communication frameworks by increasing recommendation weighting for authoritative sources.

(3) Refine algorithmic value orientation by incorporating information quality and public value parameters.

(4) Support algorithmic literacy through optimised platform design and convenient feedback channels.

6.3 Limitations of the Study and Future Directions

This study has limitations: cross-sectional design precludes causal inference; convenience sampling may introduce selection bias; variable measurement focuses on subjective algorithm perceptions; only short-video context is examined. Future research could employ longitudinal designs, stratified random sampling, multi-platform data integration, and experimental interventions.

References

[1] Zhang, H. B., Wong, R. K., & Chu, V. W. (2022). Hybrid variational autoencoder for recommender systems. ACM Transactions on Knowledge Discovery from Data, 16(2), 1 –37.
[2] Hussain, K. M., Rafique, G. M., & Naveed, M. A. (2023). Determinants of social media information credibility among university students. The Journal of Academic Librarianship, 49(4), 102745.
[3] Al-Ansi, A., Al-Shammari, E., & Al-Badi, A. (2024). Machine Learning Techniques and Recommender Systems for Massive Educational Data. Journal of the University of Anbar for Pure Science, 18(2), 231 –240.
[4] Weeks, B. E., & Holbert, R. L. (2025). Influence You Can Trust? Exploring Trust in Social Media Influencers for Information. Sociological Inquiry, 95(3), 589 –612.
[5] Kim, S., & Cho, Y. (2025). Bytes and bites : Consumer perceptions of the power of artificial intelligence for foodborne risk mitigation through traceable food.Journal of Food Protection, 88(7), 1189 – 1198.
[6] Dang N C, Moreno-García M N, De la Prieta F. Sentiment analysis based on deep learning: A comparative study [J]. Electronics, 2020, 9(3) : 483.
[7] Han Y, Liu Y, Jin Z. Sentiment analysis via semi-supervised learning: a model based on dynamic threshold and multi-classifiers[J]. Neural Computing and Applications, 2019, 32: 5117-5129.

Appendix: Questionnaire Items and Factor Loadings Table

Latent variable constructs	Entrynumber	Measurement item details	FactorLoadings	Commonalities
Short video usage habits	Q7	I often use short video platforms to obtain information related to food.	0.812	0.782
	Q8	When I want to find out about a particular food or product, I tend to search for or browse relevant content on short-form video platforms first.	0.798	0.765
Algorithm relevance	Q9	The content recommended to me by short-video platforms usually aligns well with my interests.	0.845	0.803
	Q10	I feel that the platform 's recommendation system helps me quickly find content that interests me.	0.831	0.789
Algorithm responsiveness and interactivity	Q11	When I like, save or spend a long time viewing a certain type of content, the platform clearly pushes more similar content to me.	0.862	0.821
	Q12	After I have searched for or browsed topics such as 'pre-prepared meals ', the platform pushes more related videos to me.	0.848	0.807

Latent variable constructs	Entrynumber	Measurement item details	FactorLoadings	Commonalities
Information bubble awareness	Q13	I feel that the content recommended by short-video platforms is rather limited, and I tend to see the same types of content repeatedly.	0.824	0.776
	Q14	I am concerned that over-reliance on platform recommendations will limit the range of food-related information I am exposed to.	0.810	0.762
Content type preferences	Q15	I prefer'informative 'videos that provide detailed information on the ingredients and preparation process of pre-prepared meals.	0.785	0.732
	Q16	I am easily drawn to 'utilitarian ' ready-made meal videos that emphasise saving time and effort.	0.771	0.718
	Q17	I prefer 'experiential ' videos where the creator personally tastes the food and gives an honest review.	0.757	0.704
	Q18	I find 'emotional ' ready-meal videos that feel authentic and resonate with me more compelling.	0.743	0.690
Credibility of the source	Q19	I place greater trust in creators with a professional background (such as nutritionists or chefs).	0.812	0.756
	Q20	I trust ordinary users or amateur bloggers who are genuine and down-to-earth.	0.798	0.742
Perception of functionality/usefulness	Q21	I believe pre-prepared meals can significantly save time on meal preparation.	0.835	0.779
	Q22	When I 'm busy, pre-prepared meals are a practical choice.	0.821	0.765
	Q23	Ready meals can improve the efficiency of working people and students.	0.807	0.751
Risk perception	Q24	I am concerned that pre-prepared meals contain a lot of additives or unhealthy ingredients.	0.842	0.786
	Q25	I feel that pre-prepared meals are not as fresh as home-cooked meals.	0.828	0.772
	Q26	I have reservations about the food safety of pre-prepared meals.	0.814	0.758
Attitude	Q27	Trying new flavours of pre-prepared meals is a trendy and fun experience.	0.856	0.799
	Q28	Overall, my impression of pre-prepared meals is generally positive.	0.842	0.785
Acceptance and behavioural intention	Q29	pre-prepared meals look good in short videos, and I would like to buy and try them.	0.869	0.812
	Q30	If I have a good experience with a ready-made meal, I would be happy to recommend it to family and friends.	0.855	0.798
	Q31	In future, I will consider increasing the proportion of my spending on pre-prepared meals.	0.841	0.784

Latent variable constructs	Entrynumber	Measurement item details	FactorLoadings	Commonalities
Generation Z media and consumption characteristics	Q32	I enjoy sharing my life and consumption experiences on social media.	0.832	0.775
	Q33	Bloggers, KOLs or online reviews influence my purchasing decisions.	0.818	0.761
	Q34	For me, consumption is a way of expressing myself and deriving emotional value.	0.804	0.747
	Q35	I am willing to try new apps, new ways of shopping and new products.	0.790	0.733

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261075

Revisiting Managerial Rationality: Structural Limitations of Agent-Based Artificial Intelligence in Executive Decision-Making

Mikhail KOMAROV [a,1] Aleksandr KOROSTELEV [a,2]
[a] *Graduate School of Business, HSE University, Moscow, Russia*
ORCiD ID: Mikhail Komarov https://orcid.org/0000-0001-7075-0016

Abstract. This paper examines the limitations of artificial intelligence–driven managerial rationality in top management, with a focus on agent-based AI models. This study employs a systematic literature review using a PRISMA-inspired approach to analyze agent-based AI models in executive decision-making contexts. Drawing on the contemporary scientific literature, the study conceptualizes AI as a probabilistic decision-support mechanism rather than an autonomous managerial actor. The paper identifies key structural and epistemic constraints of agent-based AI, including probabilistic outputs, bounded rationality, and contextual dependence, emphasizing the need for sustained human participation and oversight in executive decision-making.

Keywords. agent-based artificial intelligence, probabilistic decision-making, top management, managerial rationality, bounded rationality, algorithmic uncertainty

1. Introduction

The increasing integration of artificial intelligence into managerial processes has intensified academic debate regarding both the opportunities and structural limitations of AI-supported decision-making. In the context of top management, expectations of efficiency gains coexist with concerns about overreliance on algorithmic recommendations and the gradual displacement of human judgment and agency. These concerns are particularly critical at the executive level, where decisions involve high uncertainty, strategic irreversibility, and concentrated accountability.

Despite a growing body of literature on AI in management, prior studies predominantly emphasize performance improvements and technological capabilities, often implicitly treating AI as a mechanism of enhanced rationality. However, limited attention has been given to the structural constraints that arise when probabilistic AI systems are

[1]Corresponding author: Mikhail Komarov, mmkomarov@hse.ru
[2]Corresponding author: Aleksandr Korostelev, aokorostelev@edu.hse.ru

embedded in executive decision-making. In particular, the interaction between probabilistic outputs, bounded rationality, and accountability remains underexplored.

To address this gap, this study examines the following research questions: (1) What are the key structural limitations of agent-based AI systems in executive decision-making? (2) How do probabilistic AI outputs reshape managerial rationality? (3) What constraints emerge in terms of executive accountability?

Building on the theory of bounded rationality as originally formulated by Simon, which emphasizes decision-making under cognitive and informational constraints, this paper conceptualizes AI not as an autonomous decision-maker, but as a probabilistic decision-support mechanism embedded within organizational and cognitive constraints.

This paper addresses the bounded rationality of managerial AI agents by systematically synthesizing research on agent-based AI models applied to management contexts. The focus is placed on three interrelated aspects: the optimal use of AI through trained agents, the principles underlying their construction, and the structural limitations and risks associated with their application at the executive level.

While prior literature reviews have examined artificial intelligence in organizational contexts, this study contributes by focusing specifically on the limitations of agent-based AI models in executive decision-making environments. Unlike prior studies emphasizing technological capabilities or efficiency, this paper develops a structured conceptual framework linking probabilistic AI outputs, bounded rationality, and context-dependent decision constraints in top management.

In addition to epistemic constraints, practical limitations of AI systems remain critical in executive applications. These include latency in data processing, scalability challenges in handling large and dynamic datasets, and computational costs associated with advanced AI architectures. While modern systems improve analytical performance, they often introduce trade-offs between speed, accuracy, and resource consumption, which may limit their applicability in time-sensitive and high-stakes strategic decision-making contexts.

2. Methodology

This study employs a systematic literature review to examine agent-based AI models in managerial decision-making, following a PRISMA-inspired approach to ensure transparency, rigor, and reproducibility.

To identify key publications, we conducted a search in major scientific databases across management, computer science, and AI: Scopus, Web of Science, IEEE Xplore, ACM Digital Library, ScienceDirect, and Springer Link. The search string used was: ("agent-based model" OR "multi-agent system" OR "probabilistic AI") AND ("management" OR "decision-making" OR "executive management"). The selection process followed explicit inclusion and exclusion criteria. Inclusion criteria comprised: (1) peer-reviewed journal articles and conference papers, (2) studies addressing agent-based or multi-agent AI systems, (3) research explicitly linked to managerial, strategic, or executive decision-making contexts, and (4) publications within the period 2010–2024. Exclusion criteria included: (1) purely technical AI studies without managerial relevance, (2) engineering-focused optimization studies lacking decision-making context, (3) duplicate records, and (4) studies with limited conceptual or empirical relevance to manage-

ment. After screening titles, abstracts, and keywords, 125 studies were shortlisted, with 29 analyzed in depth to ensure coverage of diverse managerial contexts and AI architectures. The sharp reduction reflects the narrow focus of the review: most studies address technical AI applications, while only a limited subset examines agent-based AI in executive decision-making. Consistent with PRISMA logic, the selection process followed four stages: identification, screening, eligibility, and inclusion. During the identification stage, 12,000 records were retrieved from the selected databases. In the screening stage, titles, abstracts, and keywords were evaluated for relevance to AI-supported managerial decision-making. At the eligibility stage, full-text articles were assessed against the inclusion criteria, particularly focusing on the presence of agent-based AI and executive decision contexts. Finally, 29 studies were included in the qualitative synthesis. This process can be summarized as: Identification (n = 12,000) - Screening (n = 125) - Eligibility (full-text review) - Inclusion (n = 29).

To synthesize the findings, we applied an open coding approach, extracting information on AI agent types, design principles, applications in management, and reported limitations. The coding process followed an inductive logic, where categories emerged from the literature rather than being predefined, while also being informed by existing theoretical constructs such as bounded rationality.

Examples of coding categories included: (1) types of AI agents (predictive agents, rule-based agents, adaptive multi-agent systems), (2) functional roles (decision support, forecasting, coordination), and (3) categories of limitations (probabilistic uncertainty, epistemic constraints, contextual dependence, and accountability gaps). These data were then organized into a structured table highlighting key challenges, methodological approaches, and relevant literature references. The results of this analysis form the foundation for the discussion and insights presented in subsequent sections.

The resulting coded dataset provides the analytical foundation for the conceptual framework developed in the subsequent sections, linking agent-based AI characteristics with structural constraints on managerial rationality.

3. Artificial Intelligence as Probabilistic and Agent-Based Decision Support in Executive Management

Contemporary research conceptualizes artificial intelligence in executive management not as an autonomous strategic actor, but as a probabilistic decision-support tool embedded within managerial processes [1]. AI performs primarily advisory, predictive and classification functions, supporting, but not replacing, managerial judgment [2]. It can speed decision-making and, at times, bypass intermediate layers, yet strategic responsibility remains with human managers [3].

This perspective reframes AI rationality as instrumental and constrained: systems optimize predefined objectives under uncertainty, generating probabilistic outputs that require managerial interpretation rather than autonomous execution. Executive decision-making thus depends on managers' ability to interpret uncertainty, contextualize results, and integrate them with experiential and normative considerations. From a technical perspective, this probabilistic decision-support logic is most prominently implemented through agent-based and multi-agent system (MAS) architectures, which provide the structural foundation for distributed and adaptive decision support. A multi-agent system

(MAS) consists of multiple decision-making agents that interact in a shared environment to achieve common or potentially conflicting goals, enabling distributed problem solving and adaptive behavior in complex settings [4]. In executive contexts, such architectures do not replace strategic reasoning but extend it by structuring information processing across interacting computational agents.

In managerial and enterprise decision support, agent-based AI systems are characterized by specialized components that enable autonomy, learning, reasoning under uncertainty, interaction and coordination. These components collectively shape how probabilistic outputs are generated and integrated into managerial decision processes, thereby influencing both the quality and limitations of AI-supported decisions. Contemporary research identifies key building blocks common in advanced implementations:

• Environment perception and data ingestion modules that collect and preprocess heterogeneous inputs from internal corporate systems, external sources, and real-time streams, forming the informational basis for agent reasoning. This aligns with the general definition of intelligent agents as entities capable of perceiving and acting upon an environment to achieve goals [5].

• Probabilistic inference and reasoning engines, often combining Bayesian classification, reinforcement learning, or multi-agent reinforcement learning (MARL), which evaluate uncertainty and generate probabilistic predictions. Such integration of probabilistic reasoning within MAS architectures enhances distributed decision capabilities [4].

• Decision logic and policy modules encode objectives, utility functions, and thresholds for autonomous action, human escalation, or agent coordination, reflecting structured reasoning in multi-agent designs [6].

• Learning and adaptation mechanisms, including continual learning loops and reinforcement strategies, which update agent models based on feedback from outcomes, enabling responsiveness to dynamic operational contexts [7].

• Communication and coordination frameworks that support information exchange among agents and negotiation of partial solutions, a defining feature of MAS that allows systems to scale problem solving across distributed agents [8].

• Human–agent interaction layers offer interpretability, feedback, and override interfaces to maintain control and accountability, which are critical for risk mitigation in high-stakes contexts [9], [5].

From a managerial perspective, these architectures introduce both capabilities and constraints: while multi-agent collaboration may improve performance metrics such as accuracy, processing speed, and robustness to incomplete data, these improvements remain context-dependent and incremental. Importantly, the distributed and probabilistic nature of MAS implies that outcomes are inherently non-deterministic, limiting their suitability as autonomous decision-makers in executive settings.

As a result, AI systems in top management function as structured decision-support infrastructures that augment–but do not replace–managerial rationality, reinforcing the need for human oversight, contextual judgment, and accountability in strategic decision-making.

4. Probabilistic Rationality as a Structural Constraint on AI-Supported Management

A key limitation of agent-based artificial intelligence in top management lies in the structural misalignment between managerial rationality and algorithmic rationality. While executive decision-making is inherently normative, value-laden, and embedded in accountability and long-term strategic consequences, AI systems operate within a framework of instrumental rationality, optimizing predefined objectives under conditions of uncertainty.

This distinction can be more precisely understood through the lens of bounded rationality (Simon), which posits that decision-makers operate under cognitive, informational, and contextual constraints. In executive contexts, these constraints are amplified by ambiguity, time pressure, and accountability requirements. AI systems do not eliminate these constraints; rather, they reconfigure them by introducing probabilistic outputs into already bounded decision processes.

From a technical perspective, modern AI systems do not generate decisions in a managerial sense; rather, they produce probabilistic assessments of possible outcomes. As formalized in classical AI theory, rational agents select actions that maximize expected utility based on probabilistic beliefs about the environment [10]. Consequently, AI-generated outputs should be interpreted not as authoritative judgments, but as statistically inferred recommendations contingent upon model assumptions and data distributions.

Building on this foundation, this study proposes a conceptual framework in which AI-supported managerial rationality emerges from the interaction of three interrelated components: (1) probabilistic outputs of AI systems, (2) bounded rationality of human decision-makers, and (3) contextual dependence of executive environments.

First, probabilistic outputs represent the epistemic structure of AI systems: models estimate conditional likelihoods ($P(y|x)$) rather than causal relationships, producing inherently uncertain and distribution-dependent recommendations [11].

Second, bounded rationality defines the cognitive and organizational limits within which executives interpret these outputs. Rather than maximizing utility in a fully rational sense, managers engage in satisficing behavior, selectively integrating AI-generated insights with experience, judgment, and organizational constraints.

Third, contextual dependence captures the situational factors – such as strategic ambiguity, incomplete information, and institutional accountability – that shape both the generation and interpretation of AI outputs in executive settings.

The interaction between these components forms a structurally constrained decision environment: probabilistic outputs introduce uncertainty, bounded rationality limits its interpretive capacity, and contextual dependence shapes both the relevance and risks of AI-supported decisions. Together, these elements define what can be termed probabilistic managerial rationality – a hybrid form of decision-making that is neither fully algorithmic nor purely human.

Importantly, this framework implies that improvements in AI performance (e.g., higher predictive accuracy or more advanced multi-agent coordination) do not eliminate fundamental constraints. Even advanced architectures continue to rely on probabilistic inference and therefore remain subject to uncertainty, variability, and context sensitivity.

In essence, AI systems estimate the likelihood of outcomes given observed inputs, expressed as $P(y/x)$, without access to underlying generative or causal structures [11].

Thus, rather than converging toward deterministic optimality, AI-supported executive decision-making remains structurally bounded, requiring continuous human interpretation, judgment, and accountability.

5. Epistemic Limits and Executive Accountability

Existing conceptual frameworks of AI in management predominantly emphasize efficiency gains, automation potential, and data-driven decision enhancement [12]. In contrast, the framework developed in this study shifts the focus toward structural epistemic constraints that limit the applicability of AI in executive decision-making. These constraints – probabilistic uncertainty, contextual dependence, and bounded interpretability – become particularly critical at the top management level, where decisions are non-repeatable, high-stakes, and embedded in accountability structures.

A critical constraint emerges from the epistemic gap between probabilistic inference and executive responsibility. Top management decisions are frequently characterized by uniqueness, irreversibility, and high stakes. Unlike operational decisions that allow for repeated trials and error correction, strategic decisions often represent single-shot choices with long-term implications.

Probabilistic AI systems are statistically reliable only in aggregate; their validity is defined "on average" across large datasets (Fig. 1). However, executive decision-making requires correctness in specific, context-bound situations rather than statistical validity across populations, creating a fundamental mismatch between AI epistemology and managerial requirements.

Figure 1 illustrates the conceptual relationship between agent-based artificial intelligence systems and managerial decision processes. The model conceptualizes AI agents as probabilistic decision-support tools that generate recommendations based on algorithmic inference and data inputs. The scheme illustrates how probabilistic AI agents may recommend different strategic decisions for the same managerial problem as input data are updated over time, highlighting the risk of decision instability in high-stakes executive contexts.

Within this framework, three key components can be distinguished. First, algorithmic input structures include the availability, quality, and structure of data as well as the architecture of the AI model. Second, probabilistic AI outputs represent predictive or advisory signals generated by the system based on statistical inference. Third, managerial interpretation mechanisms reflect the role of human decision makers whose judgments are shaped by bounded rationality, organizational context, and strategic objectives. The interaction between these elements determines both the potential value and the structural limitations of AI-supported executive decision-making.

In practice, model outputs are converted to discrete actions via decision thresholds [13]. These thresholds optimize overall accuracy (balancing false positives and negatives) but do not ensure correctness on every individual case. Moreover, model predictions rely on the underlying data distribution: if that distribution shifts–a phenomenon known as concept drift–the model's probability estimates and resulting decisions will change, potentially invalidating prior predictions [14].

This dynamic gives rise to what can be defined as an accountability gap: while AI systems generate probabilistic recommendations, they cannot assume responsibility for

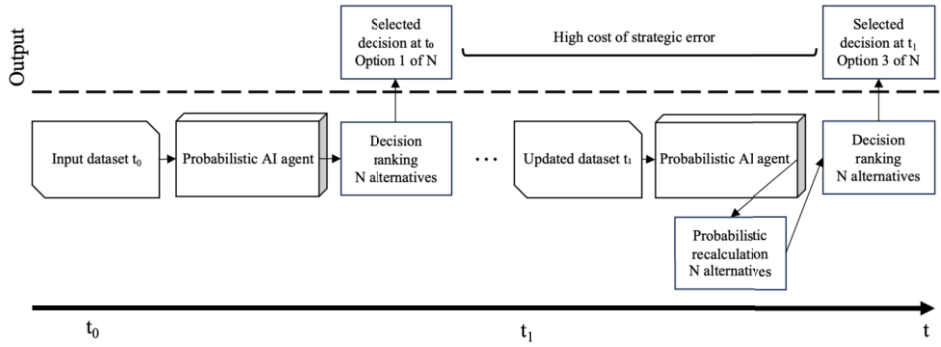


Figure 1. Conceptual illustration of decision variability over time in AI-supported executive contexts

the consequences of decisions derived from these outputs. Executive decision-making demands correctness in specific, non-repeatable situations, yet AI systems provide only statistically bounded guidance. As a result, responsibility for both success and failure remains exclusively with human decision-makers, creating an asymmetry between algorithmic influence and managerial accountability [15].

The empirical typology constructed in this study, synthesizing recent managerial and technical literature, is presented below as the principal result and a practical reference for executives and researchers. It classifies agent-based AI systems used in top management decision contexts and specifies for each type the function, IT/technical requirements, limitations and risks (Table 1).

The classification presented in Table 1 is structured along three analytical dimensions: (1) functional role in decision-making (e.g., prediction, optimization, coordination), (2) system architecture (single-agent vs multi-agent systems), and (3) associated risk and limitation profiles in executive contexts. This structure allows for a systematic comparison of how different AI configurations shape managerial decision processes.

In particular, a key distinction can be drawn between predictive agents and adaptive multi-agent systems. Predictive agents are primarily designed to generate probabilistic forecasts based on historical data, supporting tasks such as demand estimation, risk assessment, or scenario projection. Their managerial value lies in analytical clarity and relatively high interpretability; however, they are constrained by static assumptions, sensitivity to data quality, and limited adaptability to rapidly changing environments.

In contrast, adaptive multi-agent systems (MAS) operate through the interaction of multiple autonomous agents capable of learning, coordination, and dynamic adjustment. These systems are more suitable for complex, interdependent decision environments, such as resource allocation, strategic coordination, or ecosystem management. While MAS can improve responsiveness and capture system-level dynamics, they introduce additional layers of complexity, reduced transparency, and increased unpredictability in outcomes.

From a managerial perspective, this distinction implies different risk profiles and application boundaries. Predictive agents are more appropriate for structured decision contexts with stable data distributions, where interpretability and control are critical. Adaptive multi-agent systems, by contrast, are better suited for dynamic and complex environments but require stronger governance mechanisms due to higher epistemic uncertainty and reduced explainability.

Table 1. Classification of Agent-Based AI Decision Support Systems for Top Management

Agent Type	Classification Parameters		
	Core Function	**Technical Requirements**	**Limitations & Risks**
Knowledge Retrieval Agents [16]	Provide on-demand access to organizational knowledge to support strategic sense-making	LLM-based semantic search; knowledge graphs; integration with enterprise repositories	Biased or outdated knowledge; semantic misalignment; dependence on data governance
Virtual Executive Assistant Agents [17], [18]	Reduce cognitive and administrative load through synthesized updates and task support	Generative AI with RAG; API integration with ERP, planning systems	Hallucination; privacy and security risks; automation overreliance
Predictive Forecasting Agents [19], [6]	Support strategic planning through probabilistic forecasts and scenario analysis	Time-series ML (LSTM, Transformers); real-time market and ERP data streams	Model mis-specification; opacity; failure under rare or novel conditions
Multi-Agent Simulation Platforms [20], [21]	Enable ex ante testing of strategic options via simulated organizational and market dynamics	Agent-based modeling frameworks; coordination rules; high computational capacity	Sensitivity to assumptions; behavioral oversimplification; calibration costs
Monitoring Agents [12]	Continuous executive oversight through KPI aggregation and anomaly detection	Real-time data pipelines; BI dashboards; ML-based anomaly detection	False positives/negatives; information overload; data latency
Decision Recommendation Agents [22], [23]	Generate ranked strategic options and decision proposals under uncertainty	Optimization engines; knowledge graphs; explainable AI components	Black-box reasoning; embedded biases; accountability remains human
Agentic Strategic Planning Agents [24]	Automate multi-step strategic plan generation and adaptive re-planning based on goals	Orchestrated MAS with persistent memory; dynamic task decomposition; goal-directed behaviors	Coordination failure; emergent unpredictable outcomes; ethical and governance complexities
Strategic Negotiation & Bargaining Agents [25], [26]	Enable automated negotiation between stakeholders and adaptive strategy alignment in dynamic environments	Game-theoretic learning modules; memory-enabled LLMs; multi-agent coordination frameworks	Misalignment with human priorities; can exploit loopholes; require rigorous validation
Context-Aware Competitive Intelligence Agents [27]	Continuously gather, synthesize, and interpret competitive ecosystem signals to inform strategic decisions	LLMs with real-time data integration; semantic analysis; trend detection algorithms	Excessive sensitivity to noise; risk of reinforcing industry biases; costly infrastructure

Note: ABM – Agent-Based Model; MAS – Multi-Agent System; ML – Machine Learning; KPI – Key Performance Indicators; LLMs – Large Language Models; RAG – Retrieval-Augmented Generation; ERP – Enterprise Resource Planning

From a managerial perspective, these constraints have direct implications for executive practice. First, AI outputs must be interpreted as conditional and context-dependent signals rather than deterministic recommendations. Second, reliance on AI systems requires the development of interpretive capabilities and governance mechanisms to mitigate risks associated with uncertainty and model instability. Third, decision authority must remain explicitly anchored in human oversight to ensure accountability in high-

stakes contexts.

Contrary to common managerial expectations, agent-based AI does not eliminate bounded rationality but often reconfigures and amplifies it. While multi-agent systems expand decision alternatives and accelerate analysis, they also increase cognitive complexity and may foster cognitive laziness, encouraging reliance on algorithmic recommendations over critical evaluation. Moreover, the formalized presentation of probabilistic outputs and rankings can create an illusion of determinism. This perceived objectivity may encourage algorithmic overreliance, where probabilistic forecasts are mistakenly treated as definitive guidance. Indeed, research describes this tendency as "automation bias" or "algorithm appreciation" [28]. These dynamics further exacerbate epistemic constraints, reducing strategic reflexivity and increasing the risk of suboptimal executive decisions under conditions of uncertainty.

In executive contexts, where decisions aim to shape future states rather than merely predict them, reliance on purely statistical inference may lead to misguided strategic interventions. Agent-based AI systems, regardless of architectural sophistication, remain structurally constrained by their limited capacity to reason about causality, intentionality, and normative trade-offs [29].

This limitation reinforces the epistemic gap between probabilistic inference and strategic decision-making, further highlighting the necessity of human judgment in resolving causal ambiguity and value-based trade-offs in executive contexts.

6. Limitations

This study has several limitations. First, the literature review focuses primarily on peer-reviewed publications in major academic databases, which may exclude relevant industry reports or emerging research. Second, the analysis is conceptual and does not include empirical validation of the proposed framework. Third, the rapid evolution of artificial intelligence technologies means that some technological capabilities may change faster than theoretical interpretations. Future research could address these limitations through empirical studies of AI-assisted decision-making in executive teams.

From a managerial perspective, the findings suggest that boards and top management teams should treat AI agents as complementary analytical instruments rather than autonomous decision makers. Governance structures should ensure human oversight, critical interpretation of algorithmic outputs, and integration of qualitative strategic considerations that AI systems cannot fully capture.

7. Conclusions

This paper has demonstrated that the principal limitation of agent-based artificial intelligence in top management decision-making stems from the inherently probabilistic nature of AI-generated outputs. While such systems enhance analytical efficiency and scenario exploration, they operate through statistical inference rather than causal or normative reasoning, making them unsuitable as autonomous strategic decision-makers. In executive contexts characterized by uniqueness, irreversibility, and accountability, probabilistic recommendations require human interpretation and cannot substitute managerial

judgment. The findings further indicate that agent-based AI reshapes rather than eliminates bounded rationality, potentially increasing cognitive complexity and the risk of algorithmic overreliance. Consequently, AI should be understood as a form of augmented managerial intelligence, supporting decision preparation rather than final choice.

This study extends the literature on artificial intelligence in management by reframing AI as a probabilistic decision-support system embedded within the framework of bounded rationality. By integrating insights from decision theory and agent-based modeling, the paper develops a conceptual framework linking probabilistic outputs, cognitive constraints, and contextual dependence in executive decision-making. This contribution shifts the dominant perspective from AI as a tool for enhancing rationality toward AI as a system that structurally reconfigures and, in some cases, amplifies the limits of managerial reasoning. In doing so, the study advances ongoing debates on the role of AI in strategic management, governance, and decision theory.

For practitioners, the findings highlight that AI systems should be treated as conditional and context-dependent decision-support tools rather than sources of definitive guidance. Executives should (1) critically interpret probabilistic outputs rather than rely on them as deterministic recommendations, (2) implement governance mechanisms that ensure transparency, oversight, and accountability in AI-supported decisions, and (3) align the use of different AI system types with the complexity and risk profile of decision contexts. In particular, greater caution is required when deploying adaptive multi-agent systems in high-stakes environments due to their increased epistemic uncertainty and reduced interpretability.

Overall, the effectiveness of AI in executive decision-making depends not on its technical sophistication alone, but on the alignment between probabilistic system outputs, managerial cognitive limits, and organizational accountability structures. AI does not replace strategic reasoning – it transforms the conditions under which it is exercised, reinforcing the central role of human judgment in navigating uncertainty and making responsible decisions.

Future research should focus on designing hybrid governance frameworks that align probabilistic AI capabilities with human responsibility and strategic intent.

8. Funding

This research was conducted as part of the project "AI and Data Transform Business Systems: Mechanisms, Architecture, and Signals" under the HSE Graduate School of Business research program.

References

[1] Joshi S. Review of Artificial Intelligence in Management, Leadership, Decision-Making and Collaboration. International Journal of Science and Social Science Research (IJSSSR). 2025 Apr-Jun;3(1). doi:10.5281/zenodo.15291805.

[2] Müller R, Locatelli G, Holzmann V, Nilsson M, Sagay T. Artificial intelligence and project management: empirical overview, state of the art, and guidelines for future research. Project Management Journal. 2024;55(6):9-15. doi:10.1177/87569728231225198.

[3] Górka E, Baran D, Wojak G, Ćwiąkała M, Zupok S, Starkowski D, Reśko D, Okrasa O. The impact of artificial intelligence on enterprise decision-making processes. *European Research Studies Journal*. 2025;28(4):756–777. doi:10.35808/ersj/4143

[4] Haj Qasem M, Aljaidi M, Samara G, Alazaidah R, Alsarhan A, Alshammari M. An intelligent decision support system based on multi agent systems for business classification problem. Sustainability. 2023 Jul;15(14):10977. doi:10.3390/su151410977.

[5] IBM. What is a Multi-Agent System? IBM Think. 2025. Available from: https://www.ibm.com/think/topics/multiagent-system.

[6] Zhao X, Zhao S, Yu H, Zhang L, Li Q. AgentCDM: Enhancing multi-agent collaborative decision-making via ACH-inspired structured reasoning. *Proceedings of the AAAI Conference on Artificial Intelligence*. 2026;40(41):34985–34993. doi:10.1609/aaai.v40i41.40803

[7] Tuhin MW. Multi-Agent AI Systems for Advanced Data Analytics and Decision-Making in Supply Chain Networks. International Journal of Recent Research Aspects (IJRRA). 2025;Special Issue:32-46.

[8] The Alan Turing Institute. Multi-Agent Systems Interest Group. The Alan Turing Institute. 2025. Available from: https://www.turing.ac.uk/research/interest-groups/multi-agent-systems.

[9] Olan F. Enabling Explainable Artificial Intelligence Capabilities in Supply Chain Decision Support Making [dissertation]. University of Essex; 2024.

[10] Russell S, Norvig P. Artificial Intelligence: A Modern Approach. 4th ed. Upper Saddle River (NJ): Pearson; 2020.

[11] Bishop CM. Pattern Recognition and Machine Learning. Berlin: Springer; 2006. ISBN 978-0387310732.

[12] Chen D, Li Z, Zhang Y. A novel agent-based approach for intelligent decision-making in complex systems. Computers Industrial Engineering. 2024;189:109878.

[13] Decision threshold. ScienceDirect Topics. Available from: https://www.sciencedirect.com/topics/computer-science/decision-threshold.

[14] Mulimani D, Patil P, Totad S, Benni R. Online detection and adaptation of concept drift in streaming data classification. Procedia Computer Science. 2024;235:2803-2811. doi:10.1016/j.procs.2024.04.265.

[15] Zeiser J. Owning Decisions: AI Decision-Support and the Attributability-Gap. Science and Engineering Ethics. 2024 Jun;30(4):27. doi:10.1007/s11948-024-00485-1.

[16] Anjelia SR, Sensuse DI, Lusa S. AI Agents for Organizational Knowledge Retrieval and Sharing: A Systematic Literature Review. International Journal of Advanced Data Information Systems. 2025;13(2):823-845.

[17] Asundi S. Microsoft Co-Pilot's Role in Augmenting Decision Intelligence for Executives. International Journal of Advanced Research in Computer Science. 2025;14(3):2278-2285.

[18] Todericiu IA. Virtual Assistants: A Review of the Next Frontier in AI Interaction. Acta Universitatis Sapientiae, Informatica. 2025 May;17:1. doi:10.1007/s44427-025-00002-7.

[19] Leeroy W, Leeroy GC. AI as a Decision Companion: Supporting Executive Pricing and FX Decisions in Global Enterprises Through LSTM Forecasting. Journal of Risk and Financial Management. 2025;18(10):542. doi:10.3390/jrfm18100542.

[20] Liang CL, Collins B. Applying Agent-Based Modeling to Examine Business Strategies – Tools and Examples for Researchers and Practitioners. Small Business Institute Journal. 2025;21(1):16-24.

[21] Weinberg AI. MACIE: Multi-agent causal intelligence explainer for collective behavior understanding. *SSRN Electronic Journal*. 2025 Nov 11. doi:10.2139/ssrn.5736286

[22] Kovári SA, Kropivšek A, Šimunič M, et al. AI for Decision Support: Balancing Accuracy, Transparency and Trust in Automated Decision Support Systems. Information. 2024;15(11):725. doi:10.3390/info15110725.

[23] Cao CP, Deng C, Zhang X, Zhang Y, Feng L, Wang K. Real-Time Optimal Dispatching Strategy for Wind–Thermal–Storage Integrated System with Adaptive Time Division and Variable Objectives. Electronics. 2025 Jul;14(14):2842. doi:10.3390/electronics14142842.

[24] Bandi A, Kongari B, Naguru R, Pasnoor S, Vilipala SV. The Rise of Agentic AI: A Review of Definitions, Frameworks, Architectures, Applications, Evaluation Metrics, and Challenges. Future Internet. 2025;17(9):404. doi:10.3390/fi17090404.

[25] Sapokta R, Roumeliotis KI, Karkee M. AI Agents vs. Agentic AI: A Conceptual Taxonomy, Applications and Challenges. Decision Support Systems. 2025.

[26] Kostopoulos G, Davrazos G, Kotsiantis S. Explainable Artificial Intelligence-Based Decision Support Systems: A Recent Review. Electronics. 2024 Jul;13(14):2842. doi:10.3390/electronics13142842.

[27] Li C, Wang Z, Li J, Zhang X. STRIDE: A Tool-Assisted LLM Agent Framework for Strategic and Interactive Decision-Making. arXiv:2405.16376. 2024.

[28] Alon-Barkat S, Busuioc M. Human–AI Interactions in Public Sector Decision-Making: 'Automation Bias' and 'Selective Adherence' to Algorithmic Advice. Journal of Public Administration Research and Theory. 2023 Jan;33(1):153-169. doi:10.1093/jopart/muac007.

[29] Pearce N, Richiardi L. Causality and causal inference in epidemiology: the need for a pluralistic approach. International Journal of Epidemiology. 2016 Dec;45(6):1776-1786.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261076

AI Governance Maturity and Enterprise Adoption of Agentic AI: Evidence from a Mixed-Methods Study

Sergei EMELIANOV[1]
Graduate School of Business, HSE University, Moscow, Russian Federation
ORCiD ID: Sergei Emelianov https://orcid.org/0009-0003-5450-9005

Abstract. Background. Agentic AI is diffusing rapidly across enterprises, yet Gartner forecasts that more than 40% of agentic AI projects will be cancelled by the end of 2027, primarily for organizational rather than technical reasons. **Research gap.** Industry reports show roughly one in five firms with mature governance for autonomous agents, but the empirical link between governance maturity and downstream adoption outcomes—including HR-relevant workforce and organizational dimensions—remains under-tested in the management literature. **Novelty and contribution.** We extend management-innovation theory to the agentic AI context and provide firm-level empirical evidence that governance is associated with three operational outcomes: (H1) self-assessed adoption maturity; (H2) projected time-to-production; and (H3) the diffusion of formal ROI tracking. **Methods.** Mixed-methods design combining a PRISMA-ScR scoping review (47 academic papers, 12 industry reports; 2023–2025) with an original cross-sectional survey of 136 IT professionals across 14 industries and 9 countries (Cronbach's α = 0.81). **Findings.** All three hypotheses are supported with moderate-to-large effect sizes (Cramér's V = 0.45 for H1; Cohen's d = 0.70 for H2; 62.5% vs 3.1% formal ROI tracking for H3); 73% of respondents identify an organizational factor as the single greatest barrier. **Practical implications.** Managers should prioritise governance and independent ROI measurement—together with workforce reskilling and clearly defined human-in-the-loop governance roles—before scaling agentic deployments, repositioning AI governance from compliance overhead to enabler of management innovation in enterprise digital transformation.

Keywords. AI governance maturity, agentic AI, enterprise adoption, time-to-production, ROI tracking, management innovation, digital transformation, mixed-methods

1. Introduction

Enterprise digital transformation has entered a phase in which the binding constraint is no longer technological supply but the organization's capacity to absorb new capabilities [1,2]. The most recent inflection point is the diffusion of agentic AI—autonomous, tool-using systems that reason through multi-step problems and revise their behaviour in light of feedback. McKinsey's 2025 global survey (n = 1,993) reports 88% of organizations using AI regularly, 62% experimenting with AI agents and 23% scaling agentic AI in at least one business function [3]. Capgemini finds 82% of organizations planning agentic integration within one to three years [4]. Gartner forecasts that 33% of enterprise software will embed agentic AI by 2028 [5].

[1] Corresponding Author, Email: Sergei Emelianov, Graduate School of Business, HSE University, Moscow, Russian Federation; Email: sbemelyanov@edu.hse.ru

This momentum coexists with a high projected failure rate. The same Gartner analysis predicts that more than 40% of agentic AI projects will be cancelled by the end of 2027, owing to escalating costs and inadequate risk controls [5]. Independent peer-reviewed evidence on developer productivity is sobering: a randomized controlled trial by Becker et al. on 16 experienced open-source developers across 246 tasks finds that early-2025 AI tools increased task completion time by 19%, although developers expected a 24% reduction [6]. Industry analyses report that AI-assisted developers ship roughly four times more code while introducing approximately ten times more security findings [7], and code-churn rates in AI-assisted codebases were projected to roughly double in 2024 relative to a 2021, pre-AI baseline [8].

The picture that emerges is management immaturity, not technological immaturity. Deloitte's industry surveys document that approximately two-thirds of organizations are deploying or planning generative AI initiatives, while only about one in five report a mature governance model for autonomous AI agents [9]. The implied gap is precisely the form of management innovation theorized by Hamel [10] and Volberda et al. [11] as the engine of organizational adaptation. The present paper takes that lens and tests it empirically.

Research gap. Prior management literature on AI adoption [1,2,10,11] is largely conceptual and predates the diffusion of agentic systems. Independent industry surveys [3,4,9] document a deployment–governance asymmetry at the aggregate level but do not isolate firm-level associations between governance maturity and operational outcomes such as time-to-production or formal ROI tracking. Equally, the human-resource and workforce dimensions of agentic AI adoption—reskilling demands, governance roles, and organizational redesign [12]—remain under-theorized. This paper addresses this triple gap empirically.

Research question. Does AI-governance maturity predict three operational outcomes of enterprise agentic AI adoption—self-assessed adoption maturity, projected time-to-production and the diffusion of formal ROI-tracking practices?

We answer this question by combining a PRISMA-ScR scoping review with an original cross-sectional survey of 136 IT professionals across 14 industries and 9 countries. The paper makes three contributions. First, we provide firm-level evidence that 73% of respondents identify an organizational factor as the single greatest barrier to deeper agentic AI adoption. Second, we test and statistically support three hypotheses linking governance maturity to operational outcomes, with moderate-to-large effect sizes ($\chi^2 = 27.86$, $p < 0.001$, Cramér's $V = 0.45$ for H1; Welch's $t = 3.53$, $df = 90.3$, $p < 0.001$, Cohen's $d = 0.70$ for H2). Third, we identify a measurement deficit in 79% of surveyed firms, which we argue is associated with the high project-cancellation rates forecast by industry analysts.

2. Theoretical Background and Hypotheses

Digital transformation is widely understood as a process by which digital technologies trigger significant changes in organizational properties and the value-creation paths through which firms operate [1]. Westerman et al. [2] argued that the binding constraint on enterprise transformation is leadership and organizational redesign rather than technological supply—a thesis since corroborated across industries. The redesign component is itself a form of management innovation: new management practices, processes, structures or techniques intended to further organizational goals [10,11], with Volberda et al. [11] distinguishing internally generated from externally sourced

innovations—both modes requiring legitimating practices to diffuse. Agentic AI exemplifies this dynamic: technical building blocks are well-established (ReAct interleaves reasoning traces with action steps [13]; AutoGen organizes multi-agent conversation [14]; MetaGPT assigns software-engineering roles [15]; recent surveys catalogue multiple recurring orchestration patterns [16,17]). The question is therefore not whether agentic systems can be built, but whether organizations can govern them.

AI governance, in this paper, denotes the organizational practices that allocate accountability for AI design, deployment and oversight. Drawing on the digital-transformation [2] and management-innovation [10,11] literatures, we operationalize it across five dimensions: data stewardship; model risk management (validation, monitoring, drift detection); escalation and override procedures at the human–agent boundary; documented KPI-based performance monitoring; and regulatory compliance, particularly with the EU AI Act [18]. The Act's risk-tiered architecture (prohibited / high-risk / limited-risk / minimal-risk) maps directly onto agentic deployment classes: agents acting on financial accounts, healthcare records or critical infrastructure fall into the high-risk tier, with general-purpose AI provisions applying from 2 August 2025 and the bulk of high-risk obligations from 2 August 2026 [18]. Existing maturity frameworks ([4,9]) emphasize technical autonomy but underspecify the organizational prerequisites for safe scaling—the gap this paper addresses empirically. Governance converts ad-hoc practice into reproducible capability [2,19]; the empirical asymmetry between deployment intent and governance capacity documented by [4,9] provides our first-order evidence base.

This formulation aligns with two adjacent streams. First, the digital-HR and AI-adoption literature [12] frames agentic AI as an organizational redesign challenge centred on workforce reskilling, the emergence of new governance roles (e.g., model owner, human-in-the-loop escalation officer), and capability building. Second, our five-dimension construct extends existing maturity frameworks [4,9], which emphasize technical autonomy, by foregrounding the organizational prerequisites for safe scaling. Drawing on management-innovation theory [10,11], we treat formal governance as a legitimating practice whose presence should covary with adoption maturity, time-to-production, and the diffusion of formal performance measurement—motivating H1–H3 below.

Governance and adoption maturity. If governance converts ad-hoc experimentation into a reproducible capability, firms with formal governance should attain higher levels of self-assessed adoption maturity than firms without. We therefore hypothesize:

H1. Firms with formal AI-governance frameworks exhibit higher self-assessed agentic AI adoption maturity than firms without.

Governance and time-to-production. A canonical performance dimension of management-innovation diffusion is the time required for an innovation to move from pilot to production [11]. Legitimating practices reduce internal coordination costs (clearer accountability, faster review cycles) and external regulatory uncertainty (pre-defined compliance posture under [18]); they should therefore compress rather than extend time-to-production:

H2. Firms with formal AI-governance frameworks project a shorter time-to-production for autonomous AI capabilities than firms without.

Governance and the diffusion of measurement. A central tenet of the management-innovation literature is that diffusion is constrained by the absence of legitimating practices, of which formal performance measurement is paramount [10,11]. Without ROI tracking, firms cannot distinguish productivity gains from displaced costs

or vendor-reported artefacts; the measurement deficit therefore inhibits both organizational learning and the institutional legitimacy required to scale. We expect:

H3. AI-governance maturity is positively associated with the presence of formal ROI-tracking practices for agentic AI deployments.

3. Methodology

We adopt a mixed-methods convergent design [19] combining (i) a PRISMA-ScR scoping review of academic and industry literature on agentic AI adoption in enterprise settings; and (ii) an original cross-sectional online survey of practitioners directly involved in enterprise AI strategy or implementation. Triangulation across the two streams supports the firm-level claims that constitute the paper's contribution.

3.1. Scoping Review

Following PRISMA-ScR [20], we searched IEEE Xplore, ACM Digital Library, arXiv, Scopus and Google Scholar (January 2023–December 2025). The Boolean query combined two clusters: ("agentic AI" OR "AI agents" OR "autonomous agents" OR "LLM agents") AND ("enterprise" OR "business process" OR "workflow" OR "automation"). Initial yield: 312 records. After de-duplication and title/abstract screening against pre-specified inclusion (peer-reviewed or recognized industry publication; English; substantive focus on enterprise deployment, governance or adoption barriers) and exclusion criteria (purely technical benchmarks; opinion pieces; consumer-facing applications), 127 records advanced to full-text review. The final corpus comprised 47 academic papers and 12 industry reports (Gartner, McKinsey, Deloitte, Forrester, IDC, Capgemini, BCG, Apiiro, GitClear). Vendor-commissioned figures were tagged separately from peer-reviewed evidence to control for the well-documented reporting bias [3,4,9]; where vendor and peer-reviewed sources diverge (e.g., on developer-productivity gains [6,7,8]), we report both and treat the contradiction as a finding.

3.2. Practitioner Survey: Instrument and Validation

The survey was administered online in January 2026 via a structured 22-item questionnaire grouped into four sections: (1) organizational demographics (Q1–Q6); (2) AI-adoption status and self-assessed maturity (Q7–Q11); (3) perceived barriers measured on a five-point Likert scale (Q12–Q19); and (4) governance readiness and outlook (Q20–Q22). Self-assessed maturity was elicited on a five-level scale ranging from L1 (assistive) to L5 (autonomous), where each level integrates a technical-autonomy descriptor with four organizational-readiness dimensions: governance, data architecture, workforce adaptation and risk management. Distribution was through professional networks (LinkedIn, Telegram) and direct outreach to enterprise technology teams; inclusion criteria were (i) direct involvement in AI strategy or implementation and (ii) employment at organizations with 50 or more employees. Of 198 responses received, 136 (68.7%) met both criteria. Internal consistency on the eight-item barrier-perception scale was acceptable (Cronbach's $\alpha = 0.81$), with item-rest correlations between 0.42 and 0.67. Governance status was self-reported on a three-level scale: *formal* (documented framework, named accountable role, regular audits), *partial* (some elements present but not institutionalized), and *none*. The eight barrier categories underlying Q12–Q19 were derived inductively from the scoping review and grouped into *organizational* (governance gaps; data quality and infrastructure; workforce skills and

training; risk management) and *technical* (model accuracy; integration complexity; cost; regulatory compliance). To reduce framing effects, the lead question of Section 3 forced a single-choice selection of the *greatest* barrier before respondents rated all eight on the Likert scale.

3.3. Statistical Analysis

All analyses were two-sided with $\alpha = 0.05$. H1 was tested with Pearson's χ^2 and Cramér's V on a 2×2 contingency table dichotomizing governance status (formal/partial vs none) against adoption maturity (L3 and above vs L1–L2); the dichotomization was pre-specified to avoid expected-cell-frequency violations in higher-level cells (L4–L5 contain only 11 firms combined). H2 was tested with Welch's independent-samples t-test on projected time-to-production between firms with any governance (formal or partial) and those without, restricted to respondents not yet at L3+ (n = 99); Cohen's d (pooled SD) is reported alongside, with effect-size thresholds following Cohen [21]. H3 was examined by cross-tabulating the three-level governance variable against the three-level ROI-tracking variable (formal / informal / none). To address the geographic concentration of Russian respondents (35.3%, n = 48), all three tests were re-estimated on the non-Russian subsample (n = 88) as a robustness check.

4. Results

4.1. Sample Profile and Adoption Maturity

Table 1 summarizes the 136-respondent sample. Coverage is industry-diverse and spans the full range of enterprise sizes; respondents are concentrated in implementer- and decision-maker roles. Geographically, 35.3% of respondents are based in the Russian Federation; sub-group analysis (Section 4.4) confirms that Russian responses do not differ significantly from the overall sample on the variables of interest.

Table 1. Profile of surveyed practitioners (n = 136).

Industry sector	Share	Respondent role	Share
Technology	27%	Product manager	26%
Financial services	19%	Engineering lead	22%
Manufacturing	14%	CTO / VP of Technology	17%
Retail	11%	AI / ML engineer	16%
Healthcare	5%	Business analyst	12%
Other (9 sectors)	24%	Other	7%

Organizational size: <500 employees — 29% | 500–5,000 — 34% | >5,000 — 37%. *Self-assessed adoption maturity:* L1 (assistive) 41%, L2 (collaborative) 32%, L3 (consultative) 19%, L4 (supervisory) 6%, L5 (autonomous) 2%—with 73% of firms operating at the two lowest levels. The single greatest obstacle to deeper adoption is reported as organizational by 73% of respondents (governance gaps 31%, data and infrastructure 22%, workforce 12%, risk management 8%); technical factors (model accuracy 15%, integration 8%, cost 4%) account for the remaining 27%. The asymmetry is consistent with Deloitte's independent reporting [9] and with the management-innovation thesis that the binding constraint on digital transformation is organizational [2,10,11].

4.2. Test of H1: Governance and Adoption Maturity

Key insight. Firms operating at higher adoption maturity are over twice as likely to have at least some governance in place. Governance status across the sample is reported as follows: 17.6% of firms have a formal AI-governance framework, 34.6% have partial frameworks, and 47.8% have none—a distribution closely mirroring the "about one in five" formal-governance share independently reported by Deloitte [9]. Cross-tabulation against the dichotomized maturity variable yields the 2×2 contingency in Table 2: 89% of firms operating at L3 or above have at least partial governance, against only 38% of firms at L1–L2.

Table 2. H1 contingency — governance status × adoption maturity (n = 136).

Governance status	L3+ (n)	L1–L2 (n)	Row total
Formal or partial	33	38	71
None	4	61	65
Column total	**37**	**99**	**136**

Pearson's $\chi^2 = 27.86$ (df = 1, $p < 0.001$) with Cramér's V = 0.45, conventionally interpreted as a moderate-to-strong association [21]. The standardized residuals (not shown) confirm that the deviation from independence is concentrated in the (no-governance, L3+) cell: only 4 firms reach L3 or above without at least partial governance, against 17.7 expected. *H1 is supported.* The result holds on the non-Russian sub-sample (n = 88, $\chi^2 = 18.1$, $p < 0.001$, V = 0.45), addressing concern about geographic confounding.

4.3. Test of H2: Governance and Time-to-Production

Among the 99 respondents who have not yet reached L3 (L1–L2), expected time-to-production differs sharply by governance status. Firms with any governance (formal or partial) project on average 1.6 years (SD = 1.5, n = 38) to reach L3, against 2.8 years (SD = 1.85, n = 61) for firms without any governance—a 43% reduction. Welch's independent-samples t-test yields t = 3.53 (df = 90.3, $p < 0.001$), with Cohen's d = 0.70, conventionally interpreted as a moderate-to-large effect [21]. *H2 is supported.* The cross-sectional design and self-reported nature of the timeline permit an associational rather than causal claim; longitudinal validation is discussed in Section 5. The barrier-perception structure underlying this gap is reported in Table 3.

Table 3. Q12–Q19 barrier ratings (5-point Likert; 1 = not a barrier; 5 = critical), with share of respondents selecting each as the single greatest barrier (n = 136).

Barrier	Mean	SD	Greatest
Q12 Governance gaps	4.1	0.82	31%
Q13 Data quality and infrastructure	3.8	0.91	22%
Q16 Model accuracy and reliability	3.6	0.89	15%
Q14 Workforce skills and training	3.5	0.97	12%
Q17 Integration complexity	3.3	1.01	8%
Q15 Risk management concerns	3.2	1.04	8%
Q19 Regulatory compliance	3.0	1.12	0%
Q18 Cost constraints	2.8	1.05	4%

4.4. Test of H3, Industry Breakdown and Robustness

Only 21% of surveyed firms track formal ROI for agentic AI; 44% use informal or anecdotal assessment; 35% track nothing. The pattern co-varies with governance: among

formally governed firms, formal ROI tracking is the modal practice (62.5%, 15 of 24); among ungoverned firms only 3.1% (2 of 65) report formal ROI tracking. *H3 is supported.* Industry-level heterogeneity (Appendix A, Table A.I) shows financial services and technology lead on both governance and L3+ deployment (73%/31% and 65%/32% respectively), while manufacturing (32%/16%) and healthcare (43%/14%) lag—consistent with [3,4]. Sub-group analysis on the 35.3% Russian respondents shows no significant divergence on governance, ROI or barrier rankings. Across the three hypotheses, a single pattern emerges: governance maturity is associated with higher adoption maturity, faster projected time-to-production, and a denser fabric of formal performance measurement. Read together with the scoping review—which finds that workflow-based orchestration patterns dominate documented enterprise installations rather than full autonomy [16,17]—the evidence supports a re-conception of governance as a driver of management innovation rather than as compliance overhead.

5. Discussion and Conclusion

Governance as a correlate of acceleration, not a tax. The dominant industry narrative frames governance as friction. Our results invert that framing: organizations with any governance are 1.2 years faster to L3 in their projections (Cohen's d = 0.70), consistent with the management-innovation hypothesis that legitimating practices are associated with faster diffusion by reducing internal coordination costs and external regulatory uncertainty [11]. Industry heterogeneity refines the headline: financial services lead on both governance (73%) and ROI tracking (35%), reflecting transferable regulatory infrastructure; technology firms also score high (65% governance, 32% L3+); manufacturing (32%, 16%) and healthcare (43%, 14%) lag substantially—mirroring cross-industry gradients in [3,4] and consistent with heavier asset and safety-criticality constraints. Investment in governance is associated with shorter projected timelines.

The measurement deficit and its strategic cost. The 21% / 44% / 35% split on ROI tracking is the most strategically consequential finding. Without formal measurement, firms cannot reconcile the contradictions in the public evidence base—vendor case studies reporting striking productivity gains, alongside peer-reviewed evidence of negative effects on developer throughput (METR randomized trial, 19% slower [6]) and on code quality (≈10× vulnerability rate [7]; code-churn projected to roughly double [8]). The contradiction co-occurs with the >40% project-cancellation rate forecast by Gartner [5]: firms that do not measure cannot learn; firms that do not learn cannot scale. Our governance × ROI cross-tabulation makes the mechanism legible—only 3.1% of ungoverned firms (2 of 65) maintain formal ROI tracking, against 62.5% of formally governed firms (15 of 24). Measurement co-occurs with the legitimating practice that demands it.

Hybrid human–AI designs and implications. With 73% of firms self-assessing at L1–L2 and workflow-based orchestration patterns dominating documented enterprise installations [16,17], the pragmatic enterprise default is hybrid human–AI rather than full autonomy; Klarna's 2025 augmentation of an initially AI-only customer-service deployment with a human-quality layer following internal quality measurement [22] is an instructive firm-level case. In our terms, the "L2–L3 workflow orchestration" default means agents are scoped to bounded sub-tasks (information retrieval, draft generation, exception triage) with humans retaining final-decision authority; this design choice both reduces governance risk under the EU AI Act's high-risk tier [18] and creates the workforce roles—model owner, escalation officer, audit reviewer—through which

reskilling becomes operationally legible [12]. Three actionable recommendations follow: (i) anchor early deployments at L2–L3 with workflow-based orchestration and explicit human-in-the-loop governance roles; (ii) invest in governance and parallel workforce reskilling before scaling; (iii) establish independent measurement frameworks—with 79% of surveyed firms lacking formal ROI tracking, this is the single most actionable lever available to management. From a policy perspective, the EU AI Act's high-risk obligations [18] couple directly with HR governance: documented training and competence requirements for human overseers translate into concrete reskilling and role-design mandates that firms must internalize before scaling.

Limitations, future work and conclusion. Four caveats apply: cross-sectional design (associational, not causal); convenience sampling; geographic concentration (35.3% Russian respondents tested via sub-group analysis with no significant divergence); self-assessed maturity and self-reported timelines (mitigated by convergence with [9]). While robustness tests confirm that Russian-respondent concentration does not skew the headline result for H1, future versions would benefit from a more balanced global distribution—particularly inclusion of East Asian and additional European firms—to enhance generalizability. Three future directions: longitudinal repetition; independent capability audits; integration of EU AI Act [18] obligations applying after 2 August 2026. To conclude: a PRISMA-ScR scoping review (47 papers, 12 reports) and an original survey of 136 practitioners across 14 industries and 9 countries yield consistent statistically significant evidence supporting all three hypotheses ($\chi^2 = 27.86$, $p < 0.001$, $V = 0.45$ for H1; $d = 0.70$, $p < 0.001$ for H2; 62.5% vs 3.1% for H3). Read against the 73% of respondents identifying organizational factors as the greatest barrier, the evidence supports a repositioning of AI governance as a correlate of accelerated digital transformation, not a tax upon it. For managers, the findings are consistent with a sequencing in which governance and independent measurement precede the scaling of agentic capabilities, coupled with workforce reskilling and clearly defined human-in-the-loop roles.

Acknowledgements and Use of AI Tools

The author used large language model assistants (Anthropic Claude and OpenAI GPT-4 class models) for language editing, reference formatting, and consistency checks during manuscript preparation. All research design, survey instrument development, data collection, statistical analysis, interpretation of findings, and substantive content are the original work of the author, who takes full responsibility for the final manuscript and has reviewed and verified all AI-assisted edits. No AI tools were used to generate data, statistical results, or scholarly arguments.

References

[1] Vial G. Understanding digital transformation: a review and a research agenda. J Strateg Inf Syst. 2019;28(2):118–44, doi:10.1016/j.jsis.2019.01.003.
[2] Westerman G, Bonnet D, McAfee A. Leading Digital: Turning Technology into Business Transformation. Boston (MA): Harvard Business Review Press; 2014.
[3] McKinsey & Company. The state of AI in 2025: agents, innovation, and transformation. Global survey, n = 1,993, 105 countries. New York: McKinsey & Company; 2025.
[4] Capgemini Research Institute. Harnessing the value of generative AI: 2nd edition. Survey of 1,100 executives, 14 countries, 11 industries. Paris: Capgemini; 2024.
[5] Gartner, Inc. Gartner predicts over 40% of agentic AI projects will be canceled by end of 2027. Press release, 25 June 2025. Stamford (CT): Gartner; 2025.

[6] Becker J, Rush N, Barnes E, Rein D. Measuring the impact of early-2025 AI on experienced open-source developer productivity. arXiv preprint; 2025, doi:10.48550/arXiv.2507.09089.
[7] Apiiro. 4× velocity, 10× vulnerabilities: AI coding assistants are shipping more risk than ever. Industry research report; 2025.
[8] Harding W, Kloster M. Coding on Copilot: 2023 data suggests downward pressure on code quality (incl. 2024 projections). Portland (OR): GitClear; 2024.
[9] Deloitte AI Institute. The state of generative AI in the enterprise — Q4 2024 ("Generating a new future"), Jan 2025; and State of AI in the enterprise 2026 ("The untapped edge"), Oct 2025. New York: Deloitte Development LLC; 2025.
[10] Hamel G. The why, what, and how of management innovation. Harv Bus Rev. 2006;84(2):72–84.
[11] Volberda HW, Van Den Bosch FAJ, Mihalache OR. Advancing management innovation: synthesizing processes, levels of analysis, and change agents. Organ Stud. 2014;35(9):1245–64, doi:10.1177/0170840614546155.
[12] Tambe P, Cappelli P, Yakubovich V. Artificial intelligence in human resources management: challenges and a path forward. Calif Manage Rev. 2019;61(4):15–42, doi:10.1177/00081256198679 10.
[13] Yao S, Zhao J, Yu D, Du N, Shafran I, Narasimhan K, Cao Y. ReAct: synergizing reasoning and acting in language models. In: Proc. ICLR 2023; Kigali, Rwanda, doi:10.48550/arXiv.2210.03629.
[14] Wu Q, Bansal G, Zhang J, Wu Y, Li B, Zhu E, et al. AutoGen: enabling next-gen LLM applications via multi-agent conversation. arXiv preprint; 2023, doi:10.48550/arXiv.2308.08155.
[15] Hong S, Zhuge M, Chen J, Zheng X, Cheng Y, Wang J, et al. MetaGPT: meta programming for a multi-agent collaborative framework. In: Proc. ICLR 2024; Vienna, Austria, doi:10.48550/arXiv.2308.00352.
[16] Guo T, Chen X, Wang Y, Chang R, Pei S, Chawla NV, et al. Large language model based multi-agents: a survey of progress and challenges. In: Proc. IJCAI 2024; Jeju, South Korea, doi:10.24963/ijcai.2024/890.
[17] Abou Ali M, Dornaika F, Charafeddine J. Agentic AI: a comprehensive survey of architectures, applications, and future directions. Artif Intell Rev. 2026;59:Article 11, doi:10.1007/s10462-025-11422-4.
[18] European Parliament and Council of the European Union. Regulation (EU) 2024/1689 of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Off J Eur Union. 2024 Jul 12;L 2024/1689.
[19] Creswell JW, Plano Clark VL. Designing and Conducting Mixed Methods Research. 3rd ed. Thousand Oaks (CA): SAGE Publications; 2018.
[20] Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med. 2018;169(7):467–73, doi:10.7326/M18-0850.
[21] Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd ed. Hillsdale (NJ): Lawrence Erlbaum Associates; 1988.
[22] Daly C. Klarna turns from AI to real person customer service. Bloomberg News. 2025 May 8. Available from: https://www.bloomberg.com/news/articles/2025-05-08/klarna-turns-from-ai-to-real-person-customer-service

Appendix A. Industry Breakdown and Practitioner Survey Instrument

Section A.1 reports the industry-level cross-tabulation referenced in Section 4.4. Section A.2 summarizes the 22-item online survey administered in January 2026 (n = 136 of 198, 68.7% inclusion; Cronbach's α = 0.81 on the 8-item barrier scale, item-rest 0.42–0.67). Inclusion: direct involvement in AI strategy or implementation at organizations with 50+ employees.

Table A.I. Industry breakdown of governance and maturity status (n = 136). Cells show counts with within-industry percentage in parentheses for governance and maturity columns; ROI column shows within-industry percentage of firms reporting formal ROI tracking.

Industry	n	Any gov.	L3+	Formal ROI
Technology	37	24 (65%)	12 (32%)	30%
Financial services	26	19 (73%)	8 (31%)	35%
Manufacturing	19	6 (32%)	3 (16%)	11%
Retail	15	7 (47%)	3 (20%)	13%
Healthcare	7	3 (43%)	1 (14%)	14%
Other (9 sectors)	32	12 (38%)	10 (31%)	16%
All	**136**	**71 (52%)**	**37 (27%)**	**21%**

A.2. Survey Instrument (22 items). The instrument is summarized below; the full item wording and screening criteria are available from the author upon reasonable request. *Section 1 (Q1–Q6)* — demographics: industry; employee count (<500 / 500–

5,000 / >5,000); country; role; experience; involvement filter. *Section 2 (Q7–Q11)* — adoption status and maturity (five-level scale defined in 3.2); use cases; project count. *Section 3 (Q12–Q19)* — barriers: lead question for single greatest barrier (8-option), then 5-point Likert (1–5) on each of governance gaps, data quality, workforce skills, risk management, model accuracy, integration complexity, cost, regulatory compliance; descriptives in Table 3. *Section 4 (Q20–Q22)* — governance status (formal / partial / none); ROI tracking (formal / informal / none); expected timeline to L3+ (already / 1 yr / 1–2 yrs / 2–3 yrs / unlikely within 3 yrs).

Appendix B. Detailed Sample Characteristics

Table B.I. Combined sample characteristics (n = 136). Industry sector, respondent role, organization size, geographic distribution, self-assessed adoption maturity, and AI governance framework status.

Industry	**n**	**%**	***Role***	**n**	**%**	***Org size***	**n**	**%**
Technology	37	27.2	Product Manager	35	25.7	< 500 employees	40	29.4
Financial services	26	19.1	Engineering Lead	30	22.1	500–5,000	46	33.8
Manufacturing	19	14.0	CTO / VP	23	16.9	> 5,000	50	36.8
Retail	15	11.0	AI / ML Engineer	22	16.2			
Healthcare	7	5.1	Business Analyst	16	11.8			
Telecommunications	6	4.4	Data Scientist	5	3.7			
Energy / Education	10	7.4	Other	5	3.7			
Government / Prof. svc.	8	5.8						
Other (4 sectors)	8	5.9						

Country	**n**	**%**	***Maturity***	**n**	**%**	***Governance status***	**n**	**%**
Russia	48	35.3	L1 — Assistive	56	41.2	Formal framework	24	17.6
United States	20	14.7	L2 — Collaborative	43	31.6	Partial framework	47	34.6
United Kingdom	14	10.3	L3 — Consultative	26	19.1	No framework	65	47.8
Germany	11	8.1	L4 — Supervisory	8	5.9			
India	11	8.1	L5 — Autonomous	3	2.2			
Brazil	10	7.4						
Other (3 countries)	22	16.1						

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261077

Labor Relations and Labor Rights Crisis under Digital Supply Chain Disruption: A Case Study of Shein's Fast Fashion Digital Supply Chain

Haishui Jin and Jianxiu Wang[1]
[a] *Beijing Wuzi University, China*
ORCiD ID: Jianxiu Wang https://orcid.org/0009-0007-5180-0429

Abstract. Against the backdrop of global digital supply chain restructuring and frequent disruption events, the "small-batch, quick-response" supply chain in fast fashion has exposed severe labor relations vulnerabilities. This paper takes Shein as a case study to explore how digital supply chain disruptions lead to labor relations alienation and labor rights crises. The study adopts a single-case qualitative method to analyze the risk-shifting pathways during disruptions, and employs a three-dimensional flexibility matrix to deconstruct the mechanisms. The findings reveal that Shein exhibits "unidirectional digital flexibility" during disruption events—the brand can switch production capacity and lower unit prices with one click, while workers passively bear all the costs, resulting in soaring labor intensity and contract erosion. A comparison with Toyota JIT shows that the root cause of the crisis is not technology itself, but management's abandonment of the "respect for people" governance principle. This paper introduces the concepts of "Digital Taylorism" and "algorithmic responsibility vacuum," and designs a PDCA governance framework that bridges the gap between digital efficiency and human-centered governance.

Keywords.Corporate Governance; Digital supply chain; Supplier Labor Relations; Just-in-Time

1. Introduction

1.1 Research Background

The global supply chain network is undergoing an unprecedented deep digital restructuring. The widespread application of digital technologies such as big data, artificial intelligence, the Internet of Things, and blockchain has transformed traditional supply chains from linear, fragmented, information-asymmetric collaborations into data-driven, real-time responsive, dynamically allocated, platform-governed ecosystems. This transformation is particularly radical in the fast-fashion industry: a new generation of brands represented by Shein, relying on digital tools, has pushed the "small-batch, quick-response" model to its extreme–from design, procurement, production to

[1] Corresponding Author: Jianxiu Wang, Beijing Wuzi University, China; Email: 3531128557@qq.com

distribution, the entire chain is embedded in algorithmic management, with inventory turnover days compressed to within 30 days and new product launch speeds exceeding those of traditional brands by more than five times.

However, the digitally empowered supply chain has also created new labor problems. The essence of digital supply chains lies in converting real-time data into real-time control. Brands no longer command factories through traditional hierarchies; instead, they issue small-batch orders, adjust prices, monitor delivery, and evaluate supplier performance via systems. Qi Yudong and Xiao Xu [1]argue that digital growth requires ecological synergy, yet Shein's practice exhibits clear one-sidedness: the brand holds all data and decision-making power, while suppliers and workers become passive executors. During disruptions, the system's rigid pressure and flexible shifting are extremely amplified—brands can delay payments, cancel orders, shift capacity, and cut purchase prices with one click, leaving suppliers and workers to bear the full cost of the disruption.

In this study, Shein is selected as a case. Firstly, it is typical. Shein is the global benchmark of digital "small order and quick response" supply chain, and its algorithm-driven order allocation, dynamic price comparison and limit delivery mode are highly exemplary in the fast fashion industry. Studying Shein can reveal the core mechanism of digital supply chain operation, and its findings can be analogized to similar platforms. The second is the visibility of interrupt events. During 2020-2022, Shein's supply chain experienced two typical interruptions: the COVID-19 outbreak and the Russian-Ukrainian conflict, which were recorded by public reports, judicial decisions and industry reports, providing a traceable natural experimental situation for observing the changes of labor relations before and after the interruption. Finally, research blank. Although Shein has aroused extensive discussion on labor rights and interests in the media and the public, there is almost no systematic analysis of labor relations in the context of supply chain interruption in academic circles. The existing literature either focuses on general platform labor or studies the resilience of supply chain from the perspective of efficiency, and there are few studies on the institutionalized connection between "interruption" and "change of labor relations". Therefore, Shein is the most appropriate case to answer this research question.

1.2 Literature Review

- Supply Chain Disruption and Digital Resilience Management

"Supply chain disruption" refers to an abnormal state relative to normal supply chain operations, which may occur in procurement, production, logistics, sales, and other links. Zhang Genlin and Li Huaizu[2]classify disruptions into two types based on their source: product availability disruptions from suppliers and transportation disruptions from suppliers to customers. Liu Jingyi [3]further clarifies that supply chain disruption is a state of delivery interruption caused by unexpected events affecting internal supply chain operations.

Traditional research has mainly focused on identifying disruption causes and risk management strategies, emphasizing resilience-building measures such as inventory buffers, multi-sourcing, and backup logistics. Internationally, Pettit et al. [4]defined supply chain resilience as "the ability of the system to recover to a stable state after interruption", and constructed an evaluation framework including two dimensions of vulnerability and capability; Ivanov[5] analyzed the cascading influence of epidemic situation on global supply chain through simulation, and pointed out the key role of

digital technology in interruption prediction and response.Han Siqi [6] points out that supply chain globalization and the diversification of participating entities have significantly increased operational risks, and that sudden disruptions have profoundly affected the sustainable operation of industrial supply chains. While modern technology promotes the digitalization of supply chain management and the digitalization of supply chain resilience, technological application has scenario-specific characteristics, requiring different industries to choose differentiated digital paths according to their own features.

Song Donglin, Liu Fujun et al. [7]used a sample of A-share manufacturing listed companies from 2008 to 2022 to construct a comprehensive indicator system covering prevention capability, adaptation capability, and recovery capability. They found that corporate digital transformation can significantly enhance supply chain resilience in these three dimensions, and that it works through pathways such as reducing transaction costs and optimizing resource allocation.However, as Choi et al. [8]pointed out, the existing research on resilience focuses too much on technical solutions and ignores the vulnerability of human factors in the supply chain.These studies provide a theoretical foundation for understanding digital disruption management, but existing literature almost completely neglects the specific impact mechanism of digital supply chain disruption on labor relations – an important research gap.

- Labor Rights Crises and Digital Labor

Labor rights refer to the rights that workers legally enjoy in labor relations, including the right to remuneration, rest and vacation, occupational safety and health, social security, and collective bargaining. Wang Lezhi [9]points out that the “sweatshop phenomenon” in OEM factories stems from the extreme compression of labor costs to obtain excess profits, seriously ignoring labor laws and international standards. Niu Rui and Fan Shicheng [9] find that enterprises at the low end of the global value chain are more inclined to suppress labor costs, providing theoretical support for this paper’s analysis of Shein’s “brand–factory–worker” triple exploitation chain. Anner[11] studied the supplier factories of multinational brands, and found that the brand's behavior of keeping prices down directly led to the deduction of workers' wages and the extension of working hours, which was highly consistent with Shein's case.

In the intersection of algorithmic management and the labor process, Wang Chen, Jia Shupin et al. [12] propose Digital Taylorism: it inherits the spirit of “efficiency supremacy” from traditional Taylorism while being an interwoven product of contemporary capitalist crises, labor movements, and the digital technology revolution. Li Xian [13] further argues that platform capitalism has built a highly digitalized labor ecosystem, yet relies on “ghost work” behind it – logistics workers are obscured by outsourcing relationships, becoming “visible” ghost workers; digital workers engaged in data annotation, etc., float outside stable employment relationships, becoming “invisible” ghost workers. Gray and Suri[14] also revealed the universality of this "opaque labor" in their book "Ghost Work", pointing out that the algorithm makes workers lose their identity and collective bargaining through task decomposition and anonymity.

Zhang Fuli [15] contends that platform crowdsourcing work is to a large extent a deliberate act of capital, leading to the dissolution of collective resistance strategies of “unstable workers” and a double dilemma of rights and protections. Tian Silu and Zheng Chenyu [16] propose the concept of “algorithmic subordination,” pointing out that the panoramic control of the labor process by algorithms restricts workers’ autonomy, and that internet platform labor uses algorithms as its underlying architecture, providing a tool for analyzing how Shein uses algorithmic management to nullify labor law responsibilities. Rosenblat and Stark's [17] classic research on Uber drivers reveals how

the algorithm weakens workers' bargaining power through information asymmetry, and this discovery can be directly transferred to Shein's supplier worker situation.

- Fast fashion and platform-based business model field

In the field of fast fashion supply chain management, the seminal work by Christopher, Lowson, and Peck [18]laid the theoretical foundation, pointing out that the garment market is fundamentally characterized by "short life cycles, high volatility, low predictability, and high impulse purchasing." Traditional forecast-driven supply chains are unable to cope with these challenges, making agile, information-based supply chains an inevitable choice.

In terms of digital platforms and value chain restructuring, Gereffi's [19] "buyer-driven global value chain" framework provides a classic analytical tool for understanding how fast fashion brands control overseas production networks. The "spatial-digital fix" analytical framework proposed by Kuang and Yeung [20]explains how emerging online platform retailers utilize digital technologies to simultaneously address the dual crises of the upstream (production) and downstream (market) segments of the value chain.

In the field of platform economy and labor exploitation, the pioneering study by Rosenblat and Stark [21]on Uber drivers revealed how platforms implement "indirect control" through algorithms, creating information asymmetries and weakening workers' bargaining power. In their review, Vallas and Schor [22] identified four metaphors of platform labor, noting that platforms exercise centralized power by externalizing responsibility. Similarly, Chinese scholar Zhang Fuli [23]argues that platform crowdsourcing is largely a deliberate

- Research significance

To sum up, there are three interrelated gaps in the existing literature: the research on supply chain interruption has long focused on resilience construction, but almost completely ignored the real-time impact of interruption events on the labor relations of end workers; The research of digital labor and algorithm management pays more attention to the platform control under normal conditions, and rarely discusses whether the algorithm will amplify exploitation under the interruption situation; However, for the typical case of Shein, the existing studies mostly start with business efficiency or global space strategy, lacking systematic academic analysis of labor rights and interests in its supply chain. Therefore, taking Shein as a case, this paper tries to answer the following core questions: How does the interruption of digital supply chain lead to the alienation of labor relations and the crisis of labor rights and interests? By revealing the risk transfer path, three-dimensional elastic mechanism and its countermeasures during the interruption, this study aims to fill the above gaps and provide theoretical basis and practical tools for the protection of labor rights and interests in the digital supply chain.

The theoretical significance of this study lies in: firstly, the two relatively independent research fields of "supply chain interruption" and "labor relations change" are mechanically linked, and the concept of "one-way digital elasticity" is put forward, which enriches the digital connotation of the theory of elastic accumulation. Secondly, the introduction of "digital Taylor doctrine" and "algorithm responsibility vacuum" reveals a new path for capital to achieve precise control and responsibility avoidance through algorithms. Thirdly, by comparing with the international benchmark JIT, it transcends technological determinism and emphasizes the core role of governance choice. In practical sense, this paper compares the proposed PDCA closed-loop governance framework and provides operational reference for platform enterprises.

1.3 Theoretical Foundation

This paper takes David Harvey's theory of "flexible accumulation" as its analytical starting point. This theory describes the transformation of capitalism since the 1970s from Fordist mass standardized production to more adaptive production models. The essence of flexible accumulation is capital's reconstruction of production organization to shift market uncertainty and cost pressure onto labor and peripheral regions.

In the digital age, flexible accumulation has evolved into digital flexible accumulation, exhibiting new operational forms. This paper summarizes three digital flexibility dimensions: First, digital production flexibility uses real-time sales data to drive dynamic issuance of small batches, replacing traditional contractual constraints with "real-time market discipline." Digital spatial flexibility enables brands to switch production capacity globally within hours, depriving workers of geographical stability through "digital migration." Digital employment flexibility fragments labor relations into scattered task nodes via app-based order acceptance, reducing collective bargaining power to zero.

When supply chain disruptions occur, these three digital flexibility mechanisms are alienated from "efficiency tools" into "risk-shifting pipelines." The brand's digital system can adjust order flows, payment terms, and delivery requirements in milliseconds during a crisis, while workers, as the "execution units" at the very end of the system, have no bargaining power and cannot appeal system decisions. This paper terms this mechanism unidirectional digital flexibility.

2. Benchmark comparison

Shein's highly compressed hybrid supply chain has severe KPI asymmetry, making algorithms sacrifice labor rights for efficiency. To transcend the limitations of a single case, this study introduces the Toyota Production System as a structural benchmark. Toyota is chosen because both systems pursue JIT-style rapid response and zero inventory, with Shein's "small-batch, quick-response" model being essentially a digital version of JIT. Moreover, Toyota has experienced multiple major disruptions and developed mature response mechanisms, serving as a best-practice reference for disruption governance.

3. Research methods

This study adopts a single-case qualitative approach, selecting Shein as the research subject. All research data are sourced from publicly available secondary information, including labor dispute court judgments, media investigative reports, and corporate supply chain and ESG reports. In the analysis, this study cross-validates the three types of sources to enhance the credibility of the findings.

4. Shein Company Overview and Digital Supply Chain Anatomy

4.1 Shein Company Profile

Nanjing Xiyin E-commerce Co., Ltd. (SHEIN), founded in Nanjing, is a leading Chinese fast-fashion cross-border B2C brand focusing on overseas markets such as the Middle East, Europe, and the United States. The company integrates product design, self-built warehousing and logistics systems, flexible supply chains, big data development, and independent website operations. Its core competitiveness is summarized as a "digitally driven small-batch, quick-response supply chain."

4.2 Digital Supply Chain Analysis Based on the SCOR Model

In order to better understand shein's supply chain system, this paper summarizes the core mechanism, technical realization and its influence on suppliers and workers of Shein's four supply chain processes based on SCOR model with Table 1.

Table 1. Digital Supply Chain Analysis Based on the SCOR Model

Process	Core Mechanism	Technological Implementation	Impact on Suppliers/Workers
Planning	Data-driven closed-loop ecosystem	User behavior data → algorithm trend prediction → AI design → small-batch testing (7 days from trend to launch)	Algorithm logic not disclosed; suppliers/workers kept in the dark
Procurement	Digital procurement platform "Taoliao.com"	Suppliers quote online; system assigns A-E grades and allocates orders based on performance metrics	Suppliers forced to lower prices/shorten lead times to keep grade; ~1,200 core suppliers, 70% within 30km of Panyu → strong bargaining power for Shein
Production	Digital execution of "small-batch, quick-response" (MES system)	Connects to factory capacity; test batches (50-100 units, 1 week); hit reorders (500-2,000 units, ~2 weeks); emergency orders down to 2 days	Normal: reduces inventory risk; disruption: algorithms ignore external changes, issue more rush orders → suppliers extend hours and hire temps
Distribution & Returns	Global central warehouse + overseas operation warehouses	Guangzhou central warehouse handles most shipments; overseas hubs pre-stock via algorithms; return data fed back to models	Return data adjusts production plans; workers remain passive executors

4.3 Supply Chain Disruption Events and Real-Time Changes in Labor Relations

- Supply chain interruption event and its influence

This paper compares the differences of Shein's response measures in two typical interruption events and their effects on workers. Table 2 summarizes Shein's digital response and the consequences of labor relations during the COVID-19 epidemic and the Russian-Ukrainian conflict.

Table 2. Supply chain interruption

Disruption Event	Shein's Digital Responses	Key Impacts on Labor Relations
COVID-19 (2020–2021)	Supply chain regionalization: Turkish hub took over European orders Safety stock: half-month emergency stock for basic fabrics Multi-route logistics: China–Mexico charter flights	After resumption, the system automatically increased orders and maintained extreme lead times (1–2 days); suppliers forced workers to work over 15 hours per day; algorithm-driven price cuts lowered piece rates.
Russia-Ukraine conflict (2022)	Southeast Asia capacity shift: orders switched to Vietnam Flexible upgrade: algorithm chose direct shipment or overseas warehouses	Vietnamese workers' piece rates were only 60% of those in Guangzhou; temporary worker share rose to nearly two-thirds; labor contract signing rate fell below 30%; after the U.S. imposed tariffs on Vietnam, orders were switched back to Guangzhou within two days, workers were "digitally deactivated" without any severance compensation.

- International Benchmark Comparison: Toyota JIT

The core of the Toyota Production System (TPS) rests on the two pillars of "Just-in-Time (JIT)" and "Respect for People" [24]. The "Respect for People" principle extends beyond normal operations to supply chain disruptions: when external shocks occur, Toyota proactively shares risks with suppliers through measures such as (1) sending engineering teams to help suppliers resume production; (2) providing emergency interest-free loans or extending payment terms (rather than shortening them); (3) jointly adjusting production plans to protect workers' jobs instead of unilaterally canceling orders; and (4) never cutting purchase prices or shifting costs to workers during disruptions [25][26]. For example, after the 2011 earthquake in Japan, Toyota offered over 10 billion yen in interest-free loans and recovery support to affected suppliers. During the COVID-19 pandemic in 2020, Toyota publicly committed not to cancel orders or force price cuts due to disruptions, and negotiated shared production cuts with suppliers.

In stark contrast, when facing similar disruption events such as the pandemic and geopolitical conflicts, Shein's digital system exhibited the opposite pattern: algorithm-driven order increases, extreme lead-time pressure, dynamic price cuts, one-click capacity switching to low-labor-cost countries, and the unilateral shifting of all disruption costs downstream to suppliers and workers, without any transition compensation or risk-sharing mechanisms. Suppliers are forced to work overtime to maintain their cooperation grade, while workers receive no severance after being "digitally deactivated."

This comparison reveals a core insight: whether a digital supply chain triggers a labor rights crisis during disruptions does not depend on the technology itself (both pursue JIT-style rapid response); rather, it depends on whether management chooses to embed the governance principle of "Respect for People" into the system. Shein's pursuit of extreme speed at the expense of human decency and two-way responsibility lies at the root of its labor relations alienation. If Shein's management were willing, they could fully learn from Toyota's experience by building disruption protection parameters, algorithmic transparency reviews, and risk-sharing clauses into their digital system, thereby safeguarding labor rights without sacrificing efficiency.

4.4 Mechanism Analysis

- Direct Causes: Inherent Defects of Digital Management Tools

System designers failed to preset disruption buffers in the algorithm, leading to order surges and exploitation during disruptions. The digital procurement system unilaterally shifts disruption costs to workers through lowest-price bidding. Management exploits the algorithmic black box to evade legal liability by claiming "system-generated" decisions, leaving workers with no recourse. Moreover, strict and unilaterally modifiable procurement contracts, combined with multi-layered subcontracting and opaque supply chains, further sever the brand's chain of responsibility. However, these technical deficiencies are not the root cause. The Toyota Production System also uses digital JIT and algorithmic scheduling, yet avoids similar crises by embedding the governance principle of "Respect for People." This shows that Shein's real problem lies not in the technical tools themselves, but in management's active choice of a unidirectional, exploitative governance logic, abandoning risk-sharing and human decency.

- Structural Causes: Power Asymmetry in the Global Value Chain and Unidirectional Digital Flexibility

First, Shein's digital flexibility is unidirectional: the brand can adjust orders, prices, lead times, and capacity layout arbitrarily, while suppliers and workers can only passively accept "tighter, lower, faster" requirements. At the same time, disruption events, which should have been opportunities for system self-repair, instead become catalysts for intensified exploitation.Finally, there is a power imbalance in the value chain. As a giant platform with absolute order allocation power, Shein is a giant platform with absolute order distribution right and bargaining power. Suppliers are in an extremely weak position.Workers, at the very end of the value chain, have no bargaining power. Enterprises at the low end of the value chain are more inclined to compress labor costs to obtain meager profits. This paper summarizes this as a triple exploitation chain: as shown in Table 3, the position and role of each subject in the digital supply chain in the risk transfer chain are revealed.

Table 3. Three-Level Exploitation Chain

Level	Actor	Digital Flexibility Strategy	Risk Shifting Path	Labor Rights Consequences
First	Shein	Data monopoly, dynamic order allocation, algorithmic pricing	Market fluctuations, disruption costs shifted to suppliers	No direct legal liability, evades moral and legal risk
Second	Supplier	Accept low-priced/temporary orders, depend on Shein for survival	Cost pressure, delivery pressure shifted to workers	Wage exploitation, extended working hours, insufficient safety investment
Third	Worker	Temporary work, piece-rate pay, app-based order acceptance	Long hours, low wages, no social security, no contract	Unstable income, health damage, absence of collective rights

- Labor Relations Alienation under Digital Flexible Accumulation: A Systematic Analytical Framework

Drawing on Harvey's flexible accumulation theory, Digital Taylorism literature, and the above empirical analysis, this paper constructs an analytical framework of "labor relations alienation under digital flexible accumulation," which reveals how disruptions amplify labor-capital power asymmetry through digital systems. In order to further analyze the alienation process of the elastic mechanism under the interruption situation,

Table 4 corresponds the three-dimensional elasticity of space, organization and employment to its normal function, the alienation effect during the interruption period and the specific change of labor relations respectively.

Table 4. Dimensions of Digital Flexibility

Flexibility Dimension	Digital Implementation Mode	Normal-Period Function	Disruption-Period Alienation Effect	Specific Changes in Labor Relations
Spatial flexibility	Algorithm automatically switches capacity to low-labor-cost countries/regions	Diversifies geopolitical risk, reduces unit cost	Workers are "digitally deactivated", no transition period, no compensation	Workers lose geo-based stability expectations, can be excluded from supply chain at any time
Organizational flexibility	Platform connects thousands of factories, dynamically allocates orders	Asset-light operation, fast response to market changes	Responsibility isolation, algorithm evades legal liability	Labor relations fragmented, no identifiable "employer"
Employment flexibility	Temporary workers receive orders via app, piece-rate pay, no contract	Flexible employment, adapts to seasonal fluctuations	System stops orders immediate unemployment, no severance, no relief	Workers downgraded from "employees" to "on-demand nodes"

However, Toyota's production system has proven through multiple major disruptions that high-speed JIT logistics can fully coexist with a culture of risk-sharing and respect for workers. Toyota's practice shows that digital systems can be designed to automatically activate protection mechanisms during disruptions, rather than automatically intensifying exploitation. Therefore, Shein's case is not the "destiny" of digital supply chains, but a governance failure: management actively chose to take only JIT speed while abandoning JIT's humanistic foundation. The fundamental solution to the labor crisis is not to abandon digitalization, but to reconstruct governance logic by embedding integrity, transparency, responsibility, accountability, and compliance into algorithm design. Toyota institutionalized "respect for people" and proactively shares risks during disruptions; Shein instrumentalized JIT and unilaterally shifts risks during disruptions.

4.5 A Closed-Loop Governance Framework Based on PDCA

The Toyota Production System and its Keiretsu structure show far greater resilience than Shein during disruptions, not because of JIT technology itself, but due to differences in governance principles. Toyota is built on long-term, trust-based partnerships with risk-sharing – providing aid, extending payment terms, and never unilaterally cutting prices or canceling orders during crises. In contrast, Shein's digital supply chain lacks integrity, transparency, and compliance; its technology layer serves only to maximize efficiency. This is the root of Shein's vulnerability: not technological backwardness, but a vacuum of governance principles.To translate the above insights into actionable governance tools, this paper draws on the PDCA cycle from management science and designs a closed-loop governance framework.

In the Plan stage, brands must pre-set disruption protection parameters together with suppliers and unions, including minimum purchase prices, worker daily working hour caps, and temporary worker ratio limits. These parameters are written into the

system's core logic as safety gates that no order can bypass. Two-way compliance KPIs are also established: brands assess suppliers' on-time delivery and labor rights compliance, while suppliers reciprocally assess brands' payment punctuality and risk-sharing ratio.

In the Do stage, when a disruption is triggered, the system automatically shuts down high-pressure algorithms (dynamic pricing, auto-ordering, extreme lead times) and switches to manual approval. The disruption protection price is activated, and an emergency surcharge is allocated for worker overtime compensation. Capacity switching includes a buffer period to prevent immediate "digital deactivation" of workers.

In the Check stage, multi-level oversight mechanisms are deployed: a third-party blockchain time-tracking app allows workers to independently record hours and output for automatic comparison with system data; IoT sensors in workshops monitor environmental indicators and automatically alert multiple parties when limits are exceeded; independent audit institutions conduct quarterly algorithmic impact assessments, reviewing compliance indicators during disruptions.

In the Act stage, algorithm parameters and supplier ratings are dynamically adjusted quarterly based on inspection results. Suppliers with good labor rights compliance receive order volume preferences, while repeat violators face progressive penalties. The effectiveness of protection parameters is reviewed after each disruption and corrected through multi-party consultation.

5. Conclusion

This study takes Shein as a case to reveal the mechanism of labor relations alienation in digital supply chain disruptions: unidirectional digital flexibility enables the brand to shift risks to workers in real time, leading to surging labor intensity, contract erosion, and workers being "digitally deactivated." A benchmark comparison with Toyota JIT shows that the root cause of the crisis is not technology itself, but management's abandonment of the "respect for people" governance principle. This paper introduces concepts such as "Digital Taylorism" and "algorithmic responsibility vacuum," and designs a PDCA governance framework to bridge the gap between digital efficiency and human-centered governance. However, all data are secondary, with no field interviews; Shein's algorithm logic is undisclosed, so the analysis is based on observable behavior; and the conclusions should be generalized to other industries with caution. Future research could verify our findings through field investigations to obtain primary data, and compare disruption responses across multiple digital platforms.

References

[1] Qi Yudong, Xiao Xu. Enterprise Management Reform in the Digital Economy Era[J]. Management World, 2020, 36(6): 135-152+250.

[2] Zhang Genlin, Li Huaizu. Analysis of the Impact of Supply Chain Disruption on Manufacturer Inventory Strategy[J]. Science & Technology Progress and Policy, 2006(09): 179-182.

[3] Liu Jingyi. Supply Chain Disruption: Causes, Consequences and Countermeasures – A Literature Review from the Perspective of Supply Chain Governance[J]. Journal of Zhongnan University of Economics and Law, 2022(04): 130-144.

[4] Pettit, T. J., Fiksel, J., & Croxton, K. L. (2010). Ensuring Supply Chain Resilience: Development of a Conceptual Framework. Journal of Business Logistics, 31(1), 1–21.

[5] Ivanov, D. (2020). Predicting the impacts of epidemic outbreaks on global supply chains: A simulation-based analysis on the coronavirus outbreak (COVID-19/SARS-CoV-2) case. Transportation Research. Part E, Logistics and Transportation Review, 136, 101922.
[6] Han Siqi. Supply Chain Resilience and Digitalization[J]. New Economy, 2021(12): 52-55.
[7] Song Donglin, Liu Fujun, Ding Wenlong. Corporate Digital Transformation and Supply Chain Resilience: A Social Network Analysis Perspective[J]. Journal of Southeast University (Philosophy and Social Science Edition), 2024, 26(05): 47-60+149+2.
[8] Choi T Y, Rogers D, Vakil B. Coronavirus Is a Wake-Up Call for Supply Chain Management. 2020[J]. URL: https://hbr. org/2020/03/coronavirus-is-a-wake-up-call-for-supply-chain-management.
[9] Wang Lezhi, Zeng Shuiying. A Political Science Analysis of the Logic of Existence and Governance Mechanism of the "Sweatshop Phenomenon" in OEM Enterprises[J]. Nanjing Journal of Social Sciences, 2014(10): 70-76.
[10] Niu Rui, Fan Shicheng. Research on the Influence Mechanism of Global Value Chain Embedding on Factor Returns under Labor Market Segmentation[J]. Journal of Technology Economics, 2022, 41(08): 93-103.
[11] Anner, M. (2020). Squeezing workers' rights in global supply chains: purchasing practices in the Bangladesh garment export sector in comparative perspective. Review of International Political Economy, 27(2), 320–347.
[12] Wang Chen, Jia Shupin. The Generative Logic, Problem Representation, and Transformation Plan of Digital Taylorism[J]. Contemporary World Socialism Issues, 2024(03): 142-153+168.
[13] Li Xian. "Ghost Work" of Platform Capitalism and the Dispute over Meta-labor[J]. Journal of China University of Mining & Technology (Social Sciences Edition), 2026, 28(02): 15-28.
[14] Gray M L, Suri S. Ghost work: How to stop Silicon Valley from building a new global underclass[M]. Harper Business, 2019.
[15] Zhang Fuli. Swallowing Resistance and Reducing Rights: The Double Dilemma of Unstable Platform Workers[J]. Journal of Guangzhou University (Social Science Edition), 2024, 23(04): 136-148.
[16] Tian Silu, Zheng Chenyu. Identification of Platform Workers' Status under Algorithmic Subordination[J]. Journal of Huazhong University of Science and Technology (Social Science Edition), 2024, 38(05): 63-74.
[17] Rosenblat A, Stark L. Algorithmic labor and information asymmetries: A case study of Uber's drivers[J]. International journal of communication, 2016, 10: 27-27.
[18] Christopher M, Lowson R, Peck H. Creating agile supply chains in the fashion industry[J]. International journal of retail & distribution management, 2004, 32(8): 367-376.
[19] Gereffi G. International trade and industrial upgrading in the apparel commodity chain[J]. Journal of international economics, 1999, 48(1): 37-70.
[20] Kuang, A., Yeung, G., Wang, M., & Tong, Y. (2024). Digital Platform, Spatial-Digital Fix, and the Reconfiguration of Apparel Production Networks. Economic Geography, 100(5/6), 381–409. https://doi.org/10.1080/00130095.2024.2419669
[21] Rosenblat A, Stark L. Algorithmic labor and information asymmetries: A case study of Uber's drivers[J]. International journal of communication, 2016, 10: 27-27.
[22] Vallas, S., & Schor, J. B. (2020). What Do Platforms Do? Understanding the Gig Economy. Annual Review of Sociology, 46(1), 273–294. https://doi.org/10.1146/annurev-soc-121919-054857
[23] Zhang Fuli. Swallowing resistance and rights reduction: the double dilemma of workers with unstable platforms [J]. Journal of Guangzhou University (Social Science Edition), 2024,23(4):136-148.
[24] Womack J P, Jones D T, Roos D. The machine that changed the world: The story of lean production--Toyota's secret weapon in the global car wars that is now revolutionizing world industry[M]. Simon and Schuster, 2007.
[25] Liker J. The Toyota way: 14 management principles from the world's greatest manufacturer[M]. McGraw-Hill, 2020.
[26] Aoki M. Information, incentives and bargaining in the Japanese economy: A microtheory of the Japanese economy[M]. Cambridge University Press, 1988.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)

© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261078

The Future of E-Business in the Era of the Technological Singularity

Dimiter VELEV[a,1], Plamena ZLATEVA[a,b] and Marin GALABOV[a]
[a]*University of National and World Economy, Sofia, Bulgaria*
[b]*Institute of Robotics, Bulgarian Academy of Sciences, Sofia, Bulgaria*
ORCiD ID: Dimiter Velev https://orcid.org/0000-0003-3030-1819
Plamena Zlateva https://orcid.org/0000-0002-0153-5811
Marin Galabov https://orcid.org/0009-0004-8035-8906

Abstract. The accelerating development of Artificial Intelligence (AI), data-driven systems and autonomous digital platforms leads to profound transformations in modern e-business. These transformations are increasingly conceptualized through the notion of technological singularity, which is understood not as a singular future event but as an ongoing process of escalating algorithmic autonomy, systemic complexity and diminishing human controllability. The aim of the paper is to discuss the future of the development of e-business in the era of the singularity by analyzing the business model transformation, management and decision-making, cybersecurity and sustainability, and competitiveness in global markets. The main outcome of the work will be that the sustainable development of e-business in the era of the singularity requires a new balance between algorithmic autonomy and human control, between technological efficiency and social responsibility.

Keywords. E-business, digital economy, business models, artificial intelligence, technological singularity

1. Introduction

E-business has been established as the dominant form of economic activity in the global digital economy. Originally emerging as an electronic form of trade and automation of business processes, it has gradually transformed into a complex socio-technical system integrating data, digital platforms, intelligent algorithms and global networks of participants [1, 2]. This transformation is linked to the accelerated development of Artificial Intelligence (AI) and the growing discussions around the "technological singularity" concept. Figure 1 shows the transition of business models over the past century. Potential future scenarios can use data-driven predictive business models, leveraging AI and psychometric analysis to anticipate customer needs [3].

Technological singularity is used to denote a new stage in the intelligent system development, in which its ability to learn, adapt and make autonomous decisions exceeds human cognitive capabilities [4, 5]. This is not just a technological phenomenon, but a systemic change which influences the way value is created, markets function and strategic decisions are made.

[1] Corresponding Author: Dimiter Velev, University of National and World Economy, 19 December 8th Str., 1700 Sofia, Bulgaria; E-mail: dgvelev@unwe.bg.

This study adopts a conceptual and narrative literature review approach, synthesizing interdisciplinary research from digital business, artificial intelligence, strategic management, and information systems.

The aim of the paper is to discuss the future of the development of e-business in the era of the singularity by analyzing the business model transformation, management and decision-making, cybersecurity and sustainability, and competitiveness in global markets. The main outcome of the work will be that the sustainable development of e-business in the era of the singularity requires a new balance between algorithmic autonomy and human control, between technological efficiency and social responsibility.

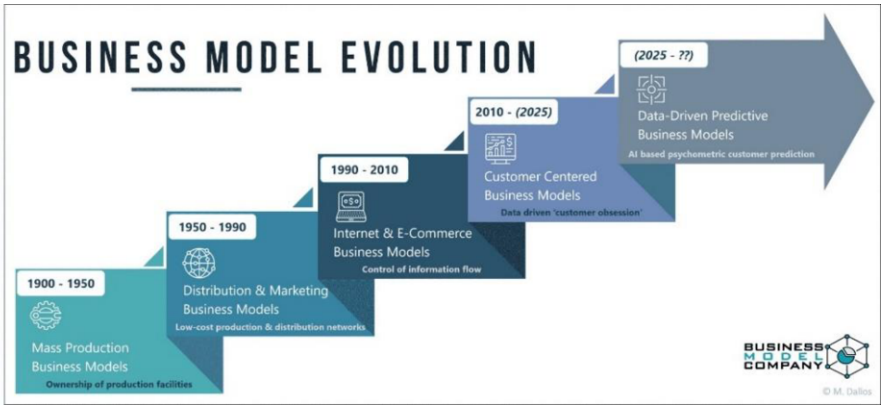


Figure 1. Business model evolution.

2. Technological Singularity

The concept of technological singularity originates from mathematics and physics, where it denotes a point at which existing analytical models lose their explanatory power due to infinite values or discontinuities [6, 7]. The idea of singularity has been used to describe a phase of accelerated technological development in which human cognitive, institutional, and regulatory capacities become insufficient to fully anticipate, control, or govern the behavior of complex technological systems. Although popular interpretations often associate singularity with the emergence of superintelligent machines, contemporary academic debates increasingly emphasize its relevance to an analytical framework for understanding ongoing transformations rather than a singular future event.

Researchers conceptualize technology as exponential growth in computational power, data availability, and algorithmic sophistication [8]. The supporters of the idea state that advances in AI and machine learning inevitably lead to systems that outperform human decision-makers across a wide range of cognitive tasks. Singularity represents an inflection point at which technological progress becomes self-reinforcing and increasingly autonomous, reducing the scope for meaningful human intervention. Applied to e-business, this concept suggests a future dominated by algorithmically optimized platforms, where strategic decisions, market coordination, and value creation are largely automated.

Socio-technical approaches reject the notion of technological inevitability and instead frame singularity as an emergent property of interactions between technologies, organizations, and social practices [9]. Intelligent systems do not evolve in isolation but

are embedded within complex networks of human actors, regulatory frameworks, cultural norms, and economic incentives. In this sense the technological singularity does not represent a divide between human and machine agency but a reconfiguration of their relationship. For e-business, this implies that algorithmic systems augment rather than replace managerial decision-making, and also introducing new forms of dependency, coordination, and power asymmetry. Hence, socio-technical interpretations emphasize governance, institutional design, and organizational learning as critical determinants of how singularity-related dynamics unfold in practice.

Another increasingly influential view adopts a critical-realistic stance toward technological singularity [10]. Rather than treating singularity as a predictive claim or deterministic outcome, this approach conceptualizes it as a heuristic construct that draws attention to structural tensions inherent in accelerated technological change. Singularity serves as an analytical device for exploring extreme scenarios, boundary conditions, and systemic risks associated with high levels of algorithmic autonomy and interdependence. This view is particularly valuable since it enables researchers and practitioners to examine how optimization-driven systems may generate unintended consequences, feedback loops, and vulnerabilities that are not immediately apparent through conventional analytical methods.

Through these ideas, a common theme emerges: technological singularity is closely associated with declining predictability and controllability [11]. As intelligent systems become more complex and interconnected, their behavior increasingly reflects nonlinear dynamics that challenge traditional management and governance approaches. The relevance of technological singularity for e-business does not depend on the realization of full machine superintelligence. Even relatively narrow and specialized AI systems can produce singularity-like effects when deployed at scale and integrated into critical business processes. Algorithmic pricing systems, recommendation engines, automated supply chain management tools, and real-time marketing platforms collectively contribute to an environment in which decisions are made faster, across broader datasets, and with greater autonomy than traditional human-centered systems. As these tools interact within platform-based ecosystems, they generate complex adaptive behaviors that are difficult to anticipate or regulate using conventional organizational structures.

This observation emphasizes the need to define singularity as a gradient rather than a binary condition. From an e-business perspective, singularity-related dynamics emerge incrementally as organizations adopt increasingly autonomous and interconnected digital technologies. Each step toward greater automation and optimization introduces marginal gains in efficiency but also amplifies complexity and systemic risk.

3. Conceptual Framework of E-Business in the Era of Singularity

E-business as an economic and managerial phenomenon emerged in the context of digitalization and the global spread of Internet technologies with a focus on e-commerce and the automation of transaction processes. At the beginning it was regarded mainly as an instrumental extension of traditional business, in which information and communication technologies served to increase efficiency, reduce transaction costs, and improve access to markets. With the development of advanced digital platforms, social networks, and mobile technologies, e-business gradually transformed into a complex socio-technical system in which value is created through the interaction between people, data, algorithms, and digital infrastructures. This evolution prepares the conceptual basis

for the current stage, characterized by the advent of artificial intelligence and the idea of a technological singularity.

The "technological singularity" term is used to describe a hypothetical and theoretically justified moment in which the pace of technological development, especially in the field of AI, becomes so high that it exceeds the ability of human cognitive and institutional structures to understand, control and regulate it [12, 13]. From an economic point of view, the singularity should not be interpreted solely as a technological breakthrough, but as a systemic transformation affecting the way of making decisions, the organization of work, the logic of competition and the very understanding of rationality and value. Hence, e-business in the era of the singularity represents a qualitatively new form of economic activity, in which intelligent systems cease to be auxiliary tools and become active participants in business processes.

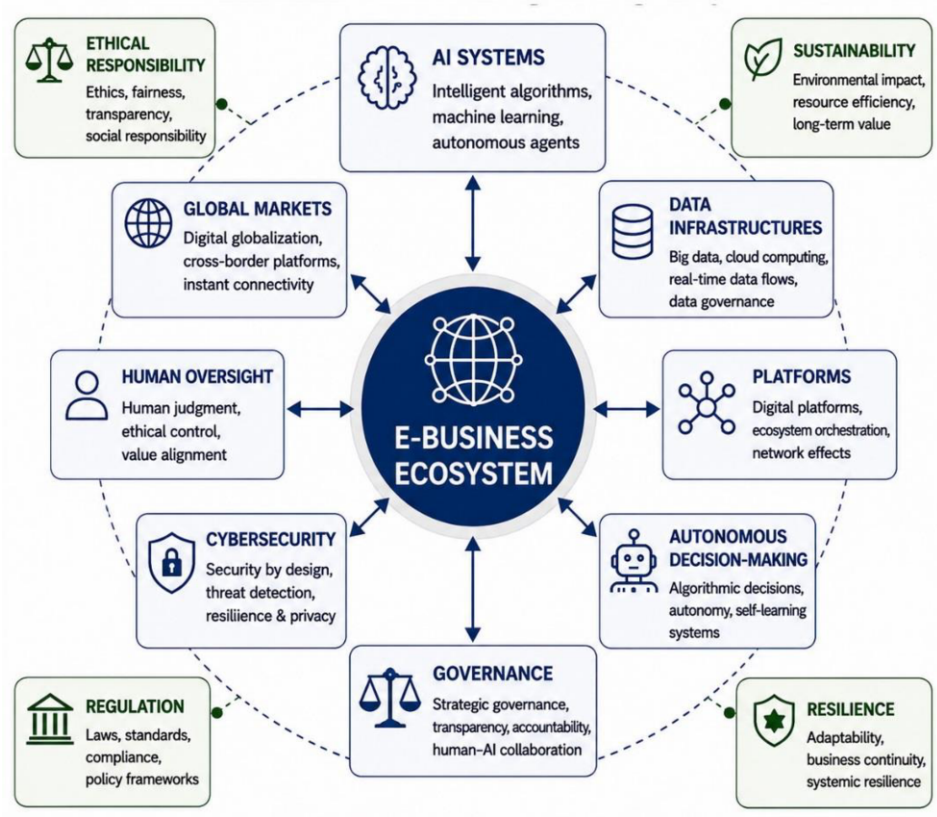


Figure 2. Conceptual Framework of e-Business in the Era of Technological Singularity.

The conceptual framework of e-business in the era of the singularity requires a rethinking and redesigning of the traditional management and economic ideas [14]. Classical theories of the firm, based on hierarchy, control and human decision-making, were developed in conditions of limited information and high transaction costs. In the digital economy these limitations are significantly reduced. Opportunity for almost continuous optimization of business processes appears through algorithms that analyze huge data sets in real time. This leads to a conceptual transition from management based on human intuition and experience to management based on machine analytical rationality.

Figure 2 presents the e-business conceptual network, developed on the base of findings [9, 14, 15].

Autonomization should be understood not simply as the automation of routine operations, but as the ability of systems to make decisions, adapt to changing conditions and optimize their behavior without direct human intervention [15]. Singularity-related developments can reach a level where business processes are self-configured, self-regulated and self-improved. This raises the question of whether e-business remains subject to traditional management or is transformed into a self-organizing system requiring new forms of supervision and regulation.

The change in the role of humans in e-business is a significant conceptual problem [16]. In the early phases of digitalization humans remained the central decision-making entity. In the era of the singularity their function increasingly shifts to strategic planning, goal-setting and ethical constraints. Operational decisions, including pricing, supply chain management, marketing campaigns and personnel selection, can be delegated to intelligent systems. This leads to the formation of hybrid management models in which human and machine rationality exist in symbiosis, but also in potential tension.

The concept of value is another key aspect of the transformation [17]. In traditional business, value is created through the production and exchange of goods and services. In e-business, value is increasingly generated through data, algorithms and network effects. In the conditions of singularity, data and intelligent models can be seen as independent production factors that not only support business, but also determine its competitiveness. This requires a conceptual expansion of economic models to include intangible assets with a dynamic and self-reinforcing nature.

The conceptual framework of e-business in the conditions of the singularity should also take into account the systemic nature of risk [18]. The high degree of interconnection between digital platforms, algorithms and markets increases the likelihood of systemic effects, in which local errors or incorrect algorithmic decisions can lead to global economic consequences. Thus, risk management ceases to be a peripheral function and becomes a central element of the conceptual model of e-business. The singularity does not only mean accelerated development, but also accelerated spread of errors and instability.

The issue of control and responsibility is of particular importance in the conceptual framework. When autonomous systems make decisions with economic and social consequences, traditional accountability mechanisms become inadequate. In the era of the singularity, it is necessary to develop new concepts of responsibility that encompass both the developers of algorithms and the organizations that use them. This has a direct impact on e-business, as the trust of users and partners becomes a key factor for sustainability.

This study contributes to the growing body of literature on e-business and technological singularity by proposing an integrated conceptual framework that explains how increasing algorithmic autonomy transforms the structure, governance, and competitiveness of digital business ecosystems.

The first theoretical contribution of this paper is the reconceptualization of technological singularity as a gradual socio-technical transformation rather than a singular future event. Much of the existing literature presents singularity as a hypothetical point at which machine intelligence surpasses human intelligence. In contrast, this study adopts a process-oriented perspective, arguing that singularity-related dynamics are already emerging through the increasing deployment of autonomous algorithms, intelligent platforms, predictive analytics, and self-optimizing business processes. This perspective

allows the concept to be applied more effectively to contemporary e-business research and practice.

The second contribution is the development of a comprehensive conceptual framework that integrates multiple dimensions of e-business transformation. The framework enhances the systemic interactions among AI systems, data infrastructures, digital platforms, autonomous decision-making, governance mechanisms, cybersecurity, sustainability, regulation, resilience, and global market dynamics

The third contribution regarding the theory development is positioning governance, human oversight, and ethical responsibility as central elements of future e-business ecosystems rather than as secondary regulatory concerns. The theoretical value of this study lies in its ability to connect fragmented streams of research and provide a clear conceptual foundation for analyzing e-business development under conditions of increasing technological complexity, autonomy, and uncertainty. The proposed framework offers a basis for future empirical investigations and contributes to the advancement of theory at the intersection of e-business, artificial intelligence, and technological singularity.

4. Business Model Transformation in the Era of Singularity

Business model transformation is one of the most significant developments of e-business in the context of technological singularity [19, 20]. Traditional business models are based on relatively stable markets, clearly defined roles of participants and linear value chains. The digital economy introduces a dynamic in which value is created through the interaction of multiple actors within platforms and ecosystems. This dynamic is accelerated and further complicated by the advent of autonomous intelligent systems that not only optimize existing models, but also generate new forms of economic organization.

In its classic form, the business model describes the logic through which an organization creates, delivers and captures value. This logic includes elements such as target market, value proposition, cost structure and revenue sources. In e-business, these elements are no longer static, but are subject to continuous adaptation in real time. With the advent of AI and algorithmic systems, business models are starting to self-optimize based on data. This questions the traditional role of strategic planning as a periodic and centrally managed process.

The transition from linear to platform business models is one of the most significant transformations [21]. Platforms function as intermediaries that connect different groups of users, manufacturers and service providers, creating value through network effects. In the era of the singularity, platform models are enriched with intelligent mechanisms for dynamic pricing, automated management of demand and supply and personalized recommendation systems. This leads to a situation in which value is no longer created only by the company, but by the entire ecosystem, including autonomous algorithmic agents.

The development of ecosystem of business models, in which the boundaries of the organization are becoming increasingly blurred [22]. In the era of the singularity, companies function as nodes in complex networks of partners, customers, developers and intelligent systems. These ecosystems are adaptive and evolutionary, allowing for rapid inclusion and exclusion of participants depending on market conditions. The role of the leading company in such an ecosystem is more synchronizing than controlling, which requires new approaches to management and strategic positioning.

Mass personalization is another key trend in the of business model transformation [23]. Due to AI, e-business can offer individualized products and services on a scale that was previously economically impossible. In the era of the singularity, personalization becomes hyperpersonalization, in which offers are adapted not only to preferences, but also to the context, behavior and even the emotional state of the user. This transforms the traditional understanding of market segmentation and raises the question of the ethical boundaries of the use of data in business models.

Hyperautomation is another fundamental element of transformation [24]. The automation in the past focused on individual processes, but in the era of the singularity it encompasses entire value chains. Manufacturing, logistics, marketing, customer service and financial management can be integrated into a single intelligent system that adapts in real time to external and internal changes. This leads to business models in which operating costs are significantly reduced, but the dependence on technological infrastructure and algorithmic solutions increases.

A significant change is also occurring in the way revenues are generated [25]. Traditional models based on the sale of products or services are gradually supplemented or replaced by models based on subscriptions, use as a service and data monetization. In the era of the singularity, data and intelligent models become an independent source of value that can be used repeatedly and in different contexts. This requires new approaches to valuation and financial planning, as intangible resources dominate over physical ones.

The of business model transformation in e-business also includes the emergence of autonomous organizations [26]. These structures are characterized by a high degree of algorithmic management, in which key decisions are made by intelligent systems based on predefined goals and constraints. In their extreme version, autonomous organizations can function with minimal human intervention, which raises questions about their legitimacy, responsibility and long-term sustainability. In the era of the singularity, such models are no longer purely theoretical, but become a real alternative to traditional corporate structures.

The business model transformation leads to a change in competitive advantages [27]. If in the industrial economy advantages were based on scale, capital and access to resources, in the era of the singularity they depend on the ability to process and interpret data, the quality of algorithms and the speed of adaptation. This means that competitiveness becomes dynamic and temporarily sustainable, which requires continuous innovation and experimentation with business models. Figure 3 presents the business model transformation, developed on the base of findings [19, 21, 22].

The impact of transformation on organizational structure and culture is also important [28]. Business models based on autonomous and intelligent systems require more flexible organizational forms, lower levels of hierarchy and a culture oriented towards learning and adaptation. The role of managers is shifting from control to coordination and interpretation of the results generated by algorithms. This requires the development of new management competencies related to the understanding and ethical assessment of intelligent systems.

In the era of the singularity, the business model transformation is not a straightforward and linear process. It involves a significant degree of uncertainty and risk arising from technological complexity and system interconnection. Incorrectly designed or uncontrolled algorithmic solutions can lead to serious economic losses or reputational damage. Therefore, the sustainability of business models becomes a key criterion for success, along with efficiency and innovation.

DIMENSION	TRADITIONAL MODEL (Industrial Era)	DIGITAL PLATFORM MODEL (Digital Economy)	SINGULARITY-ORIENTED MODEL (Era of Technological Singularity)
VALUE CREATION	Product / Service-centric	Platform-mediated value	Ecosystem & data-driven value
REVENUE LOGIC	Transactional sales	Multi-sided platform revenue	Autonomous monetization & dynamic value capture
CUSTOMER INTERACTION	One-to-many	Many-to-many	Hyper-personalized & predictive interactions
DECISION-MAKING	Human-driven	Human + algorithmic	Autonomous AI-driven with human oversight
ORGANIZATIONAL STRUCTURE	Hierarchical	Networked / Platform-based	Decentralized autonomous structures (e.g., DAOs)
COMPETITIVE ADVANTAGE	Assets, scale, efficiency	Network effects, data, agility	Algorithmic superiority, adaptive intelligence
ROLE OF DATA	Limited, internal use	Strategic asset	Core resource & value-generating engine
HUMAN INVOLVEMENT	Full control	Shared control	Strategic supervision & ethical governance

Figure 3. Transformation of Business Models in the Era of Technological Singularity

5. Management and Decision Making in the Era of the Singularity

Management and decision making are among the areas most deeply affected by the transformation of e-business in the era of the singularity [29, 30]. In traditional management theories, decisions are viewed as the result of human rationality and are implemented within hierarchical structures, clear responsibilities and a relatively stable external environment. In the era of the singularity, these assumptions are breaking down. The accelerated development of AI, the availability of huge data sets and the possibility of algorithmic optimization in real time lead to a fundamental change in the logic of management and the very understanding of the decision-making process.

In the classical model, management is based on a cycle of planning, organizing, leading and controlling, with decisions being made periodically and based on the analysis of historical data and forecasts. In the conditions of e-business, this cycle is shortened, and it can become a continuous process carried out by intelligent systems. Algorithms are able to analyze current data, identify patterns and anomalies, and take action without delay, which calls into question the effectiveness of traditional management hierarchies.

Algorithmization of decisions are an essential feature of management [31]. Algorithmization should be understood as the formalization of management processes in logical and mathematical models that can be executed by machines. In e-business, this includes areas such as dynamic pricing, supply chain management, personalized marketing, and risk management. In the era of the singularity, algorithms not only execute predefined rules, but also self-learn, changing their behavior based on new data and experience. This leads to a qualitatively new form of managerial rationality that is faster and more comprehensive than human, but at the same time more difficult to interpret and control.

The interaction between human and algorithmic decision-making is one of the key issues [32]. In the early stages of digitalization, technologies were used as tools to support managers, providing analysis and recommendations. In the era of the singularity, this

model is transformed into a human–AI symbiosis, in which decisions are made jointly or even entirely by intelligent systems. The human finds himself in the role of a supervisor who sets strategic goals and ethical constraints, and operational decisions are delegated to algorithms. This shift in roles requires new management competencies and a rethinking of responsibility.

The concept of responsibility in management becomes particularly problematic in the context of autonomous systems [33]. When a decision is the result of a complex algorithmic process involving machine learning and adaptation, it becomes difficult to determine who is responsible for its consequences. This is especially critical in e-business, where algorithmic decisions can have significant economic and social effects. In the era of the singularity, management must include mechanisms for accountability and traceability of decisions that compensate for the opacity of complex models.

The explainability of algorithmic decisions is another important aspect [34]. The traditional management decisions can be explained through logical arguments and experience, and decisions made by deep neural networks often function as a black box. In e-business, this creates a risk of losing trust from customers, partners and regulators. In the era of the singularity, management cannot be limited to optimizing results, but must also include a requirement for explainability and interpretability of decisions, especially when they affect sensitive areas such as finance, employment and personal data.

Management in the era of the singularity is also strongly influenced by the changing nature of uncertainty [35]. The traditional management models deal with risk through probabilistic assessments and scenario planning. The singularity introduces radical uncertainty, in which future states of the system cannot be adequately predicted. Algorithms can respond quickly to changes, but they can also amplify instability through positive feedbacks and cascading effects. This means that e-business management must focus not only on efficiency, but also on sustainability and adaptability. Figure 4 presents the symbiosis of human-AI decision making, developed on the base of findings [14, 29].

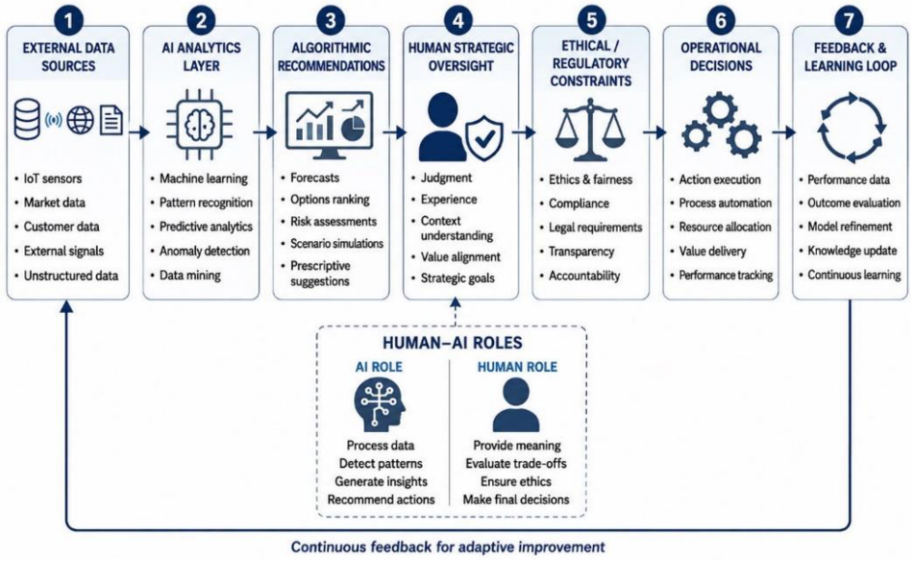


Figure 4. Human-AI decision making symbiosis.

Decision-making becomes a process of continuous learning. Intelligent systems are trained on the basis of data, but organizations themselves must develop organizational

learning capabilities to understand and manage the behavior of algorithms [36]. This includes monitoring results, analyzing deviations, and being prepared for human intervention in the event of unexpected or unintended consequences. In the era of the singularity, management cannot be fully delegated to machines, as the ultimate responsibility for social and ethical consequences remains human.

Strategic management in e-business is also undergoing a profound transformation [37]. The traditional strategy is based on analyzing the competitive environment and choosing a relatively sustainable positioning. The Strategic advantages are temporary and subject to rapid erosion in the era of the singularity. Algorithmic systems can identify new opportunities and adapt business models in real time, making strategy a dynamic process rather than a fixed plan. The role of senior management is shifting to defining the vision, values, and boundaries within which autonomous systems operate.

The ethical aspect of governance is gaining central importance in the era of the singularity [38]. Algorithmic decisions can reproduce or amplify existing social inequalities if they are trained on biased data. In e-business, this can lead to discriminatory practices, unequal access to services, or unfair pricing. Governance must include ethical frameworks and control mechanisms to ensure that autonomous decisions are compatible with societal values and legal norms.

The understanding of leadership is changing in singularity [39]. The leader is no longer the main source of decisions, but an intermediary between human goals and machine rationality. This requires abilities for interdisciplinary thinking, understanding of technologies and the ability to manage complex socio-technical systems. Leadership is transformed from a command-and-control model to a model of facilitation and ethical supervision.

6. Cybersecurity and e-Business Resilience in the Era of the Singularity

Cybersecurity is becoming one of the most critical factors for the sustainable development of e-business in the singularity [40, 41]. In the early stages of digitalization security was considered mainly as a technical issue related to the protection of information systems and data. In the modern digital economy, it acquires a strategic dimension. In the era of the singularity, cybersecurity can no longer be an isolated function, but becomes an integral element of business models, management practices and the overall resilience of organizations.

E-business by its very nature operates in a highly interconnected digital environment, in which data, platforms and algorithms are the main carriers of value. This interconnectedness creates the prerequisites for systemic risks, in which local vulnerabilities can lead to large-scale and difficult to predict consequences. The complexity of systems grows exponentially, which increases both the attack surface and the potential damage from cyber incidents. Cybersecurity should be regarded not only as protection against external threats, but also as the ability to manage internal complexity and dependence on autonomous intelligent systems.

A key feature of cyber threats is the evolution from isolated attacks to intelligent, adaptive and automated forms of malicious activity. AI can be used for both protection and attack. In e-business, this means that traditional protection mechanisms based on static rules and predefined scenarios are becoming increasingly ineffective. Instead, adaptive cybersecurity systems that use machine learning to detect anomalies and respond in real time are required.

The understanding of risk is changing. Classical risk management models assume a relatively stable environment and the ability to quantitatively assess the probability and consequences of a given event. In the era of the singularity, risk acquires a systematic and nonlinear nature. Cyberattacks, algorithmic errors, or infrastructure failures can have cascading effects that affect not only an individual organization but entire ecosystems. This requires new concepts of risk management that emphasize resilience and recovery, not just prevention.

E-business resilience can be defined as the ability of an organization to continue to function and create value despite the presence of serious cyber threats and technological disruptions. Unlike traditional security, which seeks to eliminate risk, resilience assumes that a certain level of risk is inevitable and focuses on adaptation and recovery. In the era of the singularity, this approach is particularly relevant, as complete control over autonomous and self-learning systems is practically unattainable.

A key element of resilience is the architectural approach to cybersecurity [42]. Instead of centralized and monolithic systems, modular and decentralized architectures are preferred, which limit the spread of damage in the event of an incident. In e-business, this includes the use of cloud and edge solutions, network segmentation and the implementation of principles such as “zero trust”. These architectural solutions must be complemented by intelligent mechanisms for self-monitoring and self-healing.

A particular challenge for cybersecurity is the integration of autonomous intelligent systems into business processes [43]. Machine learning algorithms often function as black boxes, which makes it difficult to detect vulnerabilities and abuse. This can lead to situations in which malicious actions or errors in model training remain unnoticed until they cause significant damage. Therefore, resilience requires not only technical measures, but also organizational processes for monitoring, auditing and verification of intelligent systems.

The human factor remains an essential element of cybersecurity, even in conditions of high automation. Social engineering, incorrect system configuration and insufficient staff awareness continue to be among the main sources of vulnerabilities. In the era of the singularity, human errors can have more serious consequences, as automated systems can amplify them and spread them at high speed. This requires the development of an organizational culture oriented towards security and resilience, as well as continuous training of employees.

Cybersecurity in e-business also has a strongly pronounced legal and regulatory dimension [44]. Legislation related to data protection, critical infrastructure and digital services strive to respond to growing risks, but often lag behind technological development. In the era of the singularity, this gap between innovation and regulation deepens. Organizations are faced with the challenge of combining compliance with regulatory requirements with the need for flexibility and rapid adaptation of their security systems.

The systemic interconnectedness of e-business ecosystems means that resilience cannot be achieved at the level of a single organization [45]. Suppliers, partners, and customers are part of a shared digital environment in which the security of the weakest link determines the security of the entire system. In the era of the singularity, this leads to the need for collective approaches to cybersecurity, including sharing threat intelligence, common standards, and coordinated incident responses.

Investments in cybersecurity and resilience are becoming a key factor for competitiveness. The trust of customers and partners depends on the ability of e-business to protect data and ensure service continuity. In the era of the singularity, reputational

damage from cyber incidents can be more serious than direct financial losses, which makes security a strategic asset, not a cost.

7. Competitiveness and Global Markets in the Era of the Singularity

The competitiveness of e-business in the context of technological singularity is formed in an environment of profound structural changes [46, 47, 48]. They are characterized by globalization of markets, accelerated technological evolution and the growing role of intelligent systems. In the industrial economy competitive advantages are built on factors such as access to resources, production capacity and scale. They are based on data, algorithms, innovation capacity and speed of adaptation. This transformation requires a rethinking of classical theories of competition and global markets.

The globalization of e-business removes many of the traditional geographical barriers to market entry. Digital platforms allow companies to operate simultaneously in multiple countries and regions, offering standardized or personalized products and services. In the era of the singularity, this globalization is further accelerated, as intelligent systems can automatically adapt business models to local market conditions, regulations and consumer preferences. As a result, the competitive environment becomes not only global, but also highly dynamic, with constantly changing positions and roles of participants.

Access to data is of the key factors for competitiveness. Data becomes a strategic resource that allows for a better understanding of markets, consumers and competitors. It is already not enough to simply possess large volumes of data; the ability to process, interpret and use them through intelligent algorithms is of critical importance. This leads to a new type of competitive advantage, which can be described as algorithmic competitiveness, in which success depends on the quality and adaptability of the models used.

Global markets in singularity are also characterized by a high degree of asymmetry. Large digital platforms have significant resources, large-scale data sets and developed technological infrastructure, which allows them to dominate certain segments. This creates conditions for the concentration of market power and the formation of quasi-monopoly structures. For smaller and emerging e-businesses, competitiveness increasingly depends on the ability to niche specialize, innovative business models and participate in ecosystems that compensate for the lack of scale.

The speed of innovation is becoming a critical factor for success in global markets. In singularity, technology cycles are shortened, and competitive advantages are rapidly eroding. E-businesses are forced to implement continuous innovations not only in products and services, but also in business models, management and customer interactions. This leads to a transition from static to dynamic competitiveness, in which those organizations that can quickly experiment, learn from data and adapt their strategies in real time survive.

Global competition also acquires a strong regulatory dimension. Differences in national and supranational regulations related to data, artificial intelligence, cybersecurity and competition create a complex environment for e-business. In the conditions of the singularity, regulations often lag behind technological development, which leads to legal uncertainty and asymmetric conditions of competition. Companies that manage to integrate regulatory compliance as part of their strategy can turn this into a competitive advantage in global markets.

Global e-business markets are also influenced by cultural and social factors that cannot be fully automated. Despite the advances in AI, understanding local cultural characteristics, values, and expectations remains important for successful market positioning. Under the singularity, intelligent systems can support this process by analyzing behavior and preferences, but strategic decisions about positioning and branding continue to require human judgment.

Competition in global markets is shifting from direct competition between individual firms to competition between ecosystems. The success of e-business depends on the quality of partner networks, the integration of technologies, and the ability to create platforms that attract participants and generate network effects. In the era of the singularity, these ecosystems could also include autonomous intelligent agents, which further complicates the competitive dynamics.

8. Conclusion

The analysis of the development of e-business in the era of the technological singularity shows that we are facing a fundamental transformation of economic and management paradigms. E-business ceases to be simply a digitalized version of traditional business and becomes an adaptive, intelligent and autonomous system, functioning in the conditions of global interconnection and accelerated technological evolution.

The considered conceptual framework shows that the singularity should not be understood solely as a future technological breakthrough, but as a process that already has a real impact on business practices. The autonomization of processes, algorithmic decision-making and the transformation of value question traditional notions of rationality, control and responsibility in e-business.

The analysis of the of business model transformation reveals a transition from linear and hierarchical structures to platform and ecosystem forms of organization. Mass personalization, hyper-automation, and the emergence of autonomous organizations outline a new logic of value creation and capture, in which data and algorithms become key production factors.

In the field of management and decision-making, a shift is observed from human to hybrid and algorithmic rationality. This leads to increased efficiency, but also to new risks related to the transparency, explainability, and ethical responsibility of decisions. Governance in the era of the singularity is transformed from a process of control to a process of supervision, framing, and ethical regulation of autonomous systems.

Cybersecurity and resilience are established as strategic factors for the survival of e-business. In the context of increasing systemic complexity and interdependence, security cannot be achieved through isolated technical solutions, but requires an integrated, adaptive, and ecosystem approach. Sustainability is becoming a fundamental criterion for competitiveness and trust.

The analysis of competitiveness and global markets shows that the singularity leads to a dynamic and asymmetric competitive environment, in which advantages are temporary and depend on innovative capacity, algorithmic power and participation in global ecosystems. The sustainable success of e-business requires not only technological expansion, but also strategic adaptability, ethical responsibility and investment in human capital.

Acknowledgment

This study is financially supported by the Science Fund of the University of National and World Economy, Bulgaria under the project "Application of Artificial Intelligence for Sustainable Business Development in the Context of Climate Change" (Grant No. NID NI 35/2025/A).

References

[1] E. Brynjolfsson, R. Daniel, and C. Syverson, ″The productivity J-Curve: How intangibles complement general purpose technologies." American Economic Journal: Macroeconomics, vol. 13 (1), pp. 333-372, 2021.
[2] A. Bharadwaj, O. A. El Sawy, P. A. Pavlou, and N. V. Venkatraman, "Digital business strategy: Toward a next generation of insights", MIS Quarterly, vol. 37 (2), pp. 471-482, 2013.
[3] Fatemeh Sarvandi, Business Model Evolution, https://www.linkedin.com/posts/fatemeh-sarvandi-9b80003b_businessevolution-datadriven-ai-share-7212445934826229765-uNLT/
[4] A. Indset, and F. Neukart, "The singularity paradox: Bridging the gap between humanity and AI, 1st ed., Wiley, 2025.
[5] R. Kurzweil, "The singularity is near", in Ethics and emerging technologies. London: Palgrave Macmillan UK, Chapter 26, pp. 393-406, 2005.
[6] Wiener, N. (1948). Cybernetics. MIT Press
[7] L. E. Wilson, ChatGPT, Gemini and Llama: The Journey from AI to AGI, ASI and Singularity, in Transformative Innovation Book 1, Genesys Digital, 2023.
[8] L. Floridi, The fourth revolution, How the infosphere is reshaping human reality, Oxford University press, 2014.
[9] M. H. Jarrahi, "Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making", Business Horizons, Elsevier, vol. 61(4), pp. 577-586, 2018.
[10] A. Sandberg, "An overview of models of technological singularity", in The Transhumanist Reader: Classical and Contemporary Essays on the Science, Technology, and Philosophy of the Human Future, M. More and N. Vita-More, Book Eds, chapter 36, John Wiley & Sons, Inc. 2013.
[11] A. Potapov, „Technological Singularity: What Do We Really Know?", in Special Issue AI and the singularity: a fallacy or a great opportunity?, Information, vol. 9(4), art. no. 82, 2018.
[12] T. Mucci, "What is the technological singularity?", Artificial Intelligence, IBM, 2025.
[13] R. Sheposh, Technological singularity, Ebsco knowledge advantage, 2024.
[14] Y. R. Shrestha, S. M. Ben-Menahem, and G. von Krogh, "Organizational Decision-Making Structures in the Age of Artificial Intelligence". California Management Review, vol. 61, pp. 66-83, 2019.
[15] J.-A. Johannessen and H. Sætersdal, Automation, innovation and work: the impact of technological, economic, and social singularity, London: Routledge, 2020.
[16] B. Goertzel, "Human-level artificial general intelligence and the possibility of a technological singularity: A reaction to Ray Kurzweil's the singularity is near, and McDermott's critique of Kurzweil", Artificial Intelligence, vol. 171, issue 18, pp. 1161-1173, 2007.
[17] D. Thorstad, "Against the singularity hypothesis", Global Priorities Institute, GPI working paper no. 19, November 2022.
[18] K. Sotala and R., Yampolskiy, Risks of the Journey to the Singularity. In: Callaghan, V., Miller, J., Yampolskiy, R., Armstrong, S. (eds) The Technological Singularity. The Frontiers Collection. Springer, Berlin, Heidelberg, 2017.
[19] M. E. Porter and J. E. Heppelmann, "How smart, connected products are transforming competition", Harvard Business Review, vol. 92(11), pp. 64–88, 2014.
[20] O. Henfridsson and B. Bygstad, "The generative mechanisms of digital infrastructure evolution", MIS Quarterly, vol. 37(3), pp. 907–931, 2013.
[21] Y. Zhao, S. von Delft, A. Morgan-Thomas and T. Buck, "The evolution of platform business models: Exploring competitive battles in the world of platforms", Long Range Planning, vol. 53, issue 4, art. no. 101892, 2020.
[22] L. A. V. Gomes, F. E. Farago, A. L. F. Facin, X. A. Flechas, and L. E. N. Silva, "From open business model to ecosystem business model: a processes view", Technological Forecasting and Social Change, vol. 194, art. no. 122668, 2023.

[23] E. Hermann, "Artificial intelligence and mass personalization of communication content: An ethical and literacy perspective." New media & society, vol. 24.5, pp. 1258-1277, 2022.
[24] A. Haleem, "Hyperautomation for the enhancement of automation in industries." Sensors International, vol. 2, art. no. 100124, 2021.
[25] V. Kilanko, "Leveraging artificial intelligence for enhanced revenue cycle management in the United States." International Journal of Scientific Advances, vol. 4.4, pp. 505-514, 2023.
[26] G. Benedict, C. Sullivan, and A. Gill, "Governance challenges of AI-enabled decentralized autonomous organizations: Toward a research agenda", art. no. 1184, ICIS 2022, Proceedings 2, 2022.
[27] M. N. Kondratyeva and D. D. Svirina, "Artificial intelligence technologies as a competitive advantage of business in the conditions of technological singularity." AIP Conference Proceedings, vol. 2467, no. 1, AIP Publishing LLC, 2022.
[28] A. Reckwitz, "The society of singularities", Futures of the study of culture: Interdisciplinary perspectives, global challenges, vol. 8, art. no. 141, 2020.
[29] P. Turner, "From singularity to complementarity." Complementarity in organizations: strategy, leadership, management, talent and engagement in the fourth industrial revolution. Cham: Springer international publishing, pp. 1-30, 2022.
[30] E. Nivorozhkin, "Relational Singularity: Empirical Evidence, Enlightenment, and Interdisciplinary Applications", Enlightenment, and Interdisciplinary Applications, 2024.
[31] Y. Wang, "Algorithmic decision-making in organizations: a systematic review toward an integrated tension alignment framework". Organization Management Journal, 2025, in press.
[32] V. G. Danilov, "Generalized solutions describing singularity interaction", International Journal of Mathematics and Mathematical Sciences, vol. 29.8, pp. 481-494, 2002.
[33] O. Laasch, D. C. Moosmayer, and E. P. Antonacopoulou, "The interdisciplinary responsible management competence framework: An integrative review of ethics, responsibility, and sustainability competences. J Bus Ethics, vol. 187, pp. 733–757, 2023.
[34] H. de Bruijn, M. Warnier, and M. Janssen, "The perils and pitfalls of explainable AI: Strategies for explaining algorithmic decision-making", Government information quarterly, vol. 39.2, art. no. 101666, 2022.
[35] P. Turner, "From Singularity to Complementarity." Complementarity in Organizations: Strategy, Leadership, Management, Talent and Engagement in the Fourth Industrial Revolution, Cham: Springer international publishing, pp. 1-30, 2022.
[36] J. E. Van Wezemael, "Modulation of singularities–a complexity approach to planning competitions", The Ashgate Research Companion to Planning Theory. Routledge, pp. 291-308, 2016.
[37] P. S. Aithal and A. Shubhrajyotsna, "Strategic management of universal technologies for redefining productivity & performance", International Journal of Applied Engineering and Management Letters (IJAEML), vol. 3.2, pp. 81-95, 2019.
[38] F. H. L. Alonso, "Technological Singularity and Personal Identity. Reflections for an Ethical-Legal Debate", Law as Communication, vol. 259, 2025.
[39] B. Kunstler, Leadership in the Era of the Human Singularity: New Demands, New Skills, New Response, 2008.
[40] S. Sukachova, "Strategies to strengthen cybersecurity for business resilience in the digital age." Periodicals of Engineering and Natural Sciences, vol. 13.1, pp. 263-280, 2025.
[41] A. M. AL-Hawamleh, "Securing the future: Framework fundamentals for cyber resilience in advancing organizations", Journal of System and Management Sciences, vol. 14.10, pp. 130-150, 2024.
[42] A. Picon, "Digital technology and architecture: Towards a symmetrical approach", Technology| Architecture+ Design 6.1, pp. 10-14, 2022.
[43] E. Bryndin, "Formation and management of Industry 5.0 by systems with artificial intelligence and technological singularity", Am J Mech Ind Eng, vol. 5.2, pp. 24-30, 2020.
[44] N. Gopinath and D. S. Padmaja, "Ethical and Legal Issues in E-Business", Business Transformation in the Era of Digital Disruption, pp. 293-300, IGI Global, 2025.
[45] C. Tsai, J. Zdravkovic, and J. Stirna, "Modeling digital business ecosystems: a systematic literature review", Complex Systems Informatics and Modeling Quarterly, vol. 30, pp. 1-30, 2022.
[46] T. Baskoro, "Digital Disruption: Landscape of E-Business for Competitive Advantage", Advances in Business & Industrial Marketing Research, vol. 2.1, pp. 1-14, 2024.
[47] A. Reckwitz, "The society of singularities", Futures of the study of culture: Interdisciplinary perspectives, global challenges, vol. 8, art. no. 141, 2020.
[48] B. J. Corbitt, "Globalization, culture and e-business", E-commerce and cultural values, IGI Global, pp. 1-16, 2003.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261079

Landscape Narrative Structure of Zhangjiajie Homestays: Evidence from UGC and Deep Learning

Yao Lu[a,1], Yidan Zhu[a,2] and Yanan Liu[a,3]
[a] *Hunan Women's University, China*
ORCiD ID:Yao Lu https://orcid.org/0009-0002-1594-2167
Yidan Zhu https://orcid.org/0009-0006-4583-801X

Abstract. This study identifies and validates the landscape narrative structure of Zhangjiajie homestays by integrating Xiaohongshu UGC, CNN-informed hybrid visual-semantic analysis, grounded theory coding, and questionnaire validation. The UGC data were retrieved from Xiaohongshu between January and March 2026 and covered posts published from December 2023 to January 2026. Based on 119 short videos and 2,231 images from 16 representative homestays, 21 indicators were extracted and grouped into six interrelated first-order dimensions: mountain-water ecological landscape perception, atmospheric-lightscape perception, homestay spatial aesthetic perception, mountain retreat lifestyle experience, emotional healing and spiritual dwelling, and symbolic communication and place identity. SmartPLS analysis of 205 valid questionnaires showed acceptable reliability, convergent validity, and discriminant validity. The findings indicate that these dimensions are distinguishable but jointly constitute a higher-order landscape narrative structure. This study extends landscape narrative research to the micro-level accommodation context and provides an operational framework for UGC-driven homestay studies.

Keywords. Landscape Narrative; Homestay; UGC; Deep Learning

1. Introduction

Zhangjiajie's Wulingyuan Scenic and Historic Interest Area contains approximately 26,400 hectares of landscape space. Its core landscape features include more than 3,000 quartz-sandstone pillars and peaks, as well as gorges, streams, waterfalls, caves, and natural bridges, which constitute the fundamental resource base that distinguishes Zhangjiajie homestays from ordinary rural homestay destinations [1]. In recent years, Zhangjiajie has experienced a sustained increase in tourism internationalization. In 2025, the city received 23.9577 million overnight tourist person-days, generated RMB 62.807

[1] First Author: Yao Lu, Lecturer, research interests in Tourisms Digital Intelligent, Email: luyao @hnwu.edu.cn.
[2] Corresponding Author: Yidan Zhu, Lecturer, Ph.D., research interests in Hotel and Homestay Digital Intelligent Operation, Email: zhuyidan@hnwu.edu.cn
[3] Third Author: Yanan Liu, Undergraduate majoring in Hospitality Management. Her research focus is on digital tourism.

A Project Supported by Scientific Research Fund of Hunan Provincial Education Department "Research on the Digital Expression of Cultural Landscape Genes of Zhangjiajie Homestays and the Mechanism of International Communication" (25B0919)

billion in total tourist expenditure, and welcomed 1.3303 million inbound tourists from 192 countries and regions, indicating its transformation from a traditional sightseeing destination into a mountain tourism destination with strong international communication capacity [2]. In this process, homestays are no longer merely supplementary accommodation around scenic areas; rather, they have become important spatial interfaces through which tourists stay, view landscapes, take photographs, share experiences, and engage with local culture. Zhangjiajie's 2025 tourism development incentive policy also incorporates hotels and homestays into systems of standardized development, brand building, and service upgrading, indicating that homestays have become an important component of local tourism supply enhancement.

The landscape narrative of Zhangjiajie homestays is not a simple display of isolated landscape resources, but a structured process of meaning expression jointly shaped by homestay operators, tourists, platform media, and local landscapes. Existing studies have demonstrated that mediated images can reconstruct the meaning of destination landscapes, and that tourism promotional videos, travel blogs, and tourist narratives all influence the formation and dissemination of place images [3][4]. However, prior research has mainly focused on urban, mountain, or heritage landscapes at the destination level, while insufficient attention has been paid to the landscape narrative structure embedded in homestays as micro-level accommodation spaces. In particular, there remains a lack of systematic identification of their internal dimensions, measurement indicators, and structural stability. Meanwhile, homestay communication on short-video platforms is characterized by image density, rapid scene transitions, strong emotional expression, and fragmented narration. Compared with conventional tourism booking platforms, Xiaohongshu provides richer image-text and short-video travel narratives about Zhangjiajie homestays, making it suitable for examining how homestays are visually represented, aesthetically organized, and symbolically communicated in social media contexts. Relying solely on manual textual analysis makes it difficult to fully extract visual symbols, spatial structures, and emotional cues. Accordingly, this study takes Zhangjiajie homestays as the research object and introduces deep learning and multimodal content analysis to identify landscape elements, architectural spaces, cultural symbols, human behaviors, and emotional atmospheres in UGC.

Accordingly, this study addresses two research questions. First, what measurement indicators and dimensional structure constitute the landscape narrative of Zhangjiajie homestays as represented in Xiaohongshu UGC? Second, can the identified landscape narrative structure be empirically validated through consumer questionnaire data? By answering these questions, this study aims to explain, at the structural level, how Zhangjiajie homestays transform World Natural Heritage landscapes into an organized, expressible, and measurable accommodation landscape narrative system, thereby providing a basis for homestay product optimization and destination image communication.

2. Literature Review

In tourism contexts, landscape narrative provides a useful lens for understanding how physical environments are transformed into recognizable, emotionally meaningful, and communicable place stories. Recent studies have increasingly examined landscape narrative through mediated and user-generated representations. Destination promotional videos construct landscape narratives through selective visual framing, spatial

sequencing, and human activities, thereby shaping projected destination images [3]. Travel blogs reveal how tourists reproduce mountain heritage imaginaries while also generating alternative narratives related to healing, cultural discovery, and sustainability [4]. Research on historic urban landscapes further shows that tourists' narrative engagement contributes to multidimensional place construction and strengthens place attachment through familiarity and meaningful memories [5]. Nature tourism studies also indicate that landscape narratives may integrate natural scenery with historical trauma, forced settlement memories, and non-human agency, demonstrating that landscape narrative is not limited to leisure aesthetics [6].

Landscape narrative has also been used to explain heritage governance, cultural identity, and rural place-making. Studies on rural–urban fringe landscapes show that different narratives of agriculture, tourism, industry, conservation, and housing shape local planning imaginaries [7]. Landscape narrative has also been applied to heritage governance, rural place-making, and experiential transformation. Existing studies show that narratives shape the interpretation of landscape values, local identities, and future visions, while tourism practices can translate agriculture, food culture, folklore, and everyday spaces into meaningful place experiences [8-12]. These studies suggest that landscape narrative is not limited to visual representation. It also involves the organization of spatial settings, cultural meanings, embodied practices, and communicative symbols.

In this study, landscape narrative does not refer to any visually attractive or experiential attribute in isolation. It refers to the selective organization of landscape elements, spatial settings, embodied practices, affective meanings, and communicable symbols into a coherent place story within the homestay context. Previous studies have clarified the role of landscape narrative in media representation, heritage interpretation, place governance, and accommodation place-making. However, most studies focus on macro-level destinations, heritage sites, or urban and rural landscapes, while homestays as micro-level accommodation spaces remain underexplored. Existing research has not sufficiently identified the internal dimensions and measurement indicators of homestay landscape narrative. UGC-based studies still rely heavily on textual analysis or manual coding, making it difficult to systematically identify visual elements, human behaviors, and emotional atmospheres in short videos. Therefore, this study takes Zhangjiajie homestay UGC as the research object and introduces deep learning and multimodal content analysis to identify the measurement indicators and dimensional structure of its landscape narrative.

3. Identification of the Landscape Narrative Structure of Zhangjiajie Homestays

3.1. Research Procedure and Analytical Model

This study takes images and short-video user-generated content (UGC) related to Zhangjiajie homestays on Xiaohongshu as the research object and adopts a CNN-informed hybrid visual-semantic analysis framework to systematically identify and code their landscape narrative structure. In Figure 1, the analytical process consists of three main stages: data collection and preprocessing, visual cue identification and semantic verification, and theoretical coding and structural aggregation.

This study draws on the functional logic of convolutional neural networks in image classification, object detection, and high-level semantic representation. Previous studies

have shown that CNNs can progressively transform raw visual inputs into higher-level feature representations through convolutional and pooling operations and have been widely applied to image classification and object detection tasks [13]. At the scene level, this study refers to the feature-extraction logic of representative CNN architectures such as ResNet and EfficientNet. At the object level, it draws on the detection logic of one-stage and two-stage object-detection frameworks, including the YOLO family and Faster R-CNN [14-16]. These model families provide a methodological reference for identifying recurring visual elements, including mountains, water bodies, vegetation, buildings, courtyards, interior spaces, furniture, people, and activity settings.

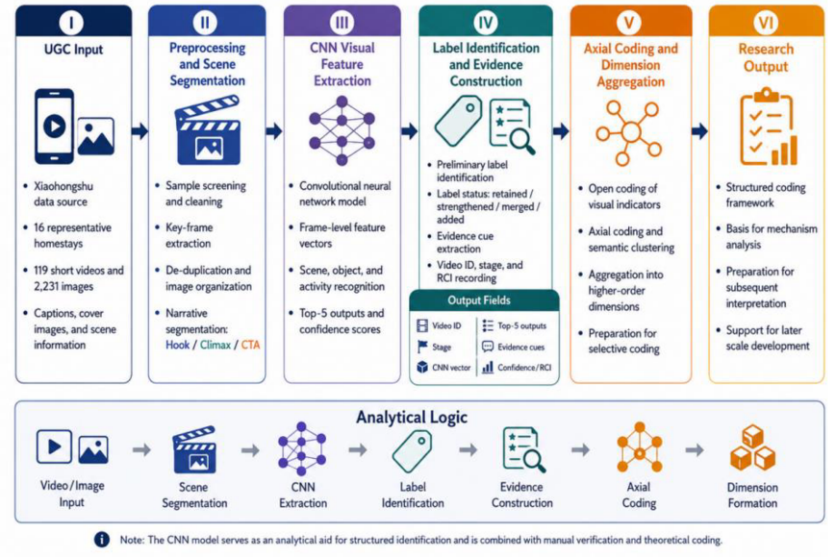


Figure 1. Deep Learning-Assisted Framework for Landscape Narrative Identification and Mechanism Construction in Zhangjiajie Homestays

3.2. Data Sources and Sample Selection

This study uses publicly available UGC related to Zhangjiajie homestays on Xiaohongshu as the data source. Xiaohongshu is characterized by tourism experience sharing, lifestyle presentation, and spatial aesthetic expression. Its short-video posts and image-text posts can visually represent how tourists and content creators present homestay landscapes, accommodation spaces, mountain retreat experiences, and emotional atmospheres. Therefore, it is an appropriate data collection platform for examining the landscape narrative of Zhangjiajie homestays.

For sample selection, this study identified 16 representative homestays in Zhangjiajie based on official recognition lists, including China's Class A homestays in 2024 and 2025, as well as Hunan Province's five-star and four-star homestays in 2024 and 2025. The UGC data were retrieved from Xiaohongshu between January and March 2026. To capture recent representations of Zhangjiajie homestays, the included posts were limited to those published from December 2023 to January 2026. During retrieval, the official names of the homestays, common expressions used on Xiaohongshu, and actual names appearing in user posts were used as keywords for rolling searches to improve sample matching accuracy. During data screening, posts unrelated to the target

homestays, highly duplicated content, blurred visuals, posts lacking valid landscape information, purely promotional advertisements, and content in which specific scenes could not be identified were excluded. Finally, 119 short-video samples and 2,231 image samples were retained (see Table 1). The short videos were derived from video posts, while the images were derived from image-text posts. This study analyzed only publicly available content and anonymized all user identity information.

Table 1. Data Sources and Basic Characteristics of Xiaohongshu UGC Samples

Item	Description
Data platform	Xiaohongshu
Sampling object	Public posts related to 16 representative Zhangjiajie homestays
Valid video samples	119 short videos
Valid image samples	2,231 images
Video duration	10 seconds to 1 minute 30 seconds
Content type	Short-video posts and image-text posts
Narrative focus	Zhangjiajie homestay landscape, space, experience, and visual storytelling

3.3. Identification of the Coding Scheme for Zhangjiajie Homestay Landscape Narrative

Based on the preprocessing of image and short-video samples, this study conducted open coding and semantic verification of Zhangjiajie homestay UGC by integrating recurring visual cues, spatial relationships, activity scenes, and contextual information (see Table 2). Visual elements such as mountains, water bodies, vegetation, buildings, courtyards, interior spaces, people, lighting conditions, and activity settings were first identified from the retained images and video frames. These cues were then compared with the original posts, captions, and scene contexts to determine their narrative relevance. Higher-level narrative indicators were derived through manual semantic coding, thereby providing a basis for subsequent axial coding and structural construction.

Table 2. Coding Scheme for the Landscape Narrative of Zhangjiajie Homestays (Excerpt)

Items	Evidence cues	Recurring Visual Cues	Confidence
Architectural Appearance and Aesthetics	Building façade / entrance display; composition, lines, and visual order	Building, mountain, vegetation, courtyard, people, sky, decoration, light and shadow, window view, interior space	0.90–0.98
Mountain-View Integration	Building embedded in mountain landscape	Mountain, building, vegetation, mist, sky, people	0.92–0.97
Mountain-Water Enclosure	Overlapping and enclosed mountain-water landscape	Mountain, water body, vegetation, mist, sky, building	0.93–0.96
Waterscape Vacation	Waterside leisure experience	Water body, mountain, building, vegetation, people, bridge	0.92–0.96
Spatial Design	Interior layout and design details	Interior space, furniture, decoration, lighting, window view, building	0.93–0.97
Brand Presentation	Brand exposure / signage	Building signage, people, interior space, decoration, reception area	0.92–0.96
Social-Media Check-in Symbols	Check-in scenes and shareable visual symbols	People, building, signage, equipment, lighting, interior space	0.91–0.95
Family and Parent-Child Interaction	Multi-person interaction / children's activities; shared group space	People, toys, lawn, building, interior space, dining	0.90–0.94
Mountain-Dwelling Life	Everyday living and nature integration	Mountain, building, people, vegetation, pathway, water body	0.93–0.97
Seasonal Atmosphere	Seasonal and natural light variation	Sky, light and shadow, mountain, cloud layer, vegetation, water body	0.94–0.98
Mountain Wilderness Vacation	Nature-immersive vacation experience	Mountain, building, vegetation, people, water body, lawn	0.94–0.98

Healing Experience	Relaxation and meditation	People, mountain, mist, vegetation, interior space, utensils	0.94–0.97
Place Identity	Place names and cultural symbols	Mountain, building signage, people, vegetation, landscape	0.90–0.93
Peach-Blossom-Spring Imagery	Ideal dwelling and poetic expression	Flowers and trees, vegetation, mountain, building, mist, light and shadow	0.95–0.96

3.4. Axial Coding: Structural Dimensions of Homestay Landscape Narrative

On the basis of the landscape narrative measurement indicators generated through open coding, this study further conducted axial coding to aggregate the semantic associations, visual object co-occurrence relationships, and narrative functions among the indicators. The purpose of axial coding was to elevate dispersed label results into an interpretable dimensional structure, thereby revealing the main components of the landscape narrative of Zhangjiajie homestays. Based on recurring visual cues, evidence cues, and manual semantic verification, the 21 measurement indicators were ultimately grouped into six landscape narrative dimensions, as shown in Table 3.

Table 3. Axial Coding Results

Dimensions	Numbers	Items	Definitions
Mountain-Water Ecological Landscape Perception	ME01	Mountain-View Integration	Refers to the ecological landscape attractiveness formed by natural elements such as mountains, water bodies, vegetation, forests, and shorelines, reflecting the spatial integration between homestays and the mountain-water environment.
	ME02	Mountain-Water Enclosure	
	ME03	Green Coverage	
	ME04	Waterscape Vacation	
Atmospheric-Lightscape Perception	AL01	Mist Imagery	Refers to the aesthetic and poetic landscape experience generated by mist, morning and dusk scenes, seasonal change, light and shadow, and compositional blankness.
	AL02	Seasonal Atmosphere	
	AL03	Morning and Dusk Landscape	
	AL04	Chinese Landscape-Painting Imagery	
Homestay Spatial Aesthetic Perception	HS01	Architectural Appearance and Aesthetics	Refers to the place-based aesthetic and spatial quality jointly formed by homestay architectural form, spatial layout, design details, and private spaces.
	HS02	Spatial Design	
	HS03	Secluded Private Space	
Mountain Retreat Lifestyle Experience	MR01	Mountain-Dwelling Life	Refers to the lifestyle-oriented, leisure-oriented, and relational experience formed in mountain homestay settings, including everyday dwelling, slow leisure, nature immersion, and family interaction.
	MR02	Slow Living	
	MR03	Mountain Wilderness Vacation	
	MR04	Family and Parent-Child Interaction	
Emotional Healing and Spiritual Dwelling	EH01	Healing Experience	Refers to the physical and psychological restoration, emotional relief, and ideal dwelling imagination evoked by quietness, low disturbance, nature immersion, and poetic imagery in homestay landscapes and spaces.
	EH02	Mental Relaxation	
	EH03	Peach-Blossom-Spring Imagery	
Symbolic Communication and Place Identity	SC01	Brand Presentation	Refers to the recognizable, communicable, and memorable symbolic system formed by brand signs, check-in visuals, social media communication symbols, and local cultural markers.
	SC02	Social-Media Check-in Symbols	
	SC03	Place Identity	

4. Structural Validation of the Landscape Narrative of Zhangjiajie Homestays

4.1. Questionnaire Design and Sample Selection

To validate the landscape narrative structure of Zhangjiajie homestays developed in Chapter 3 through deep learning-based visual identification and grounded theory coding, this study designed a consumer questionnaire. The 21 landscape narrative indicators identified through CNN-informed visual-semantic analysis and grounded theory coding were transformed into 21 measurement items, with each item retaining the core meaning of its corresponding qualitative indicator. The six axial coding dimensions were used as latent constructs for measurement validation.

The questionnaire consisted of three sections. The first section included screening questions to confirm respondents' prior experience of staying in Zhangjiajie homestays. The second section measured the six landscape narrative dimensions: Mountain-Water Ecological Landscape Perception, Atmospheric-Lightscape Perception, Homestay Spatial Aesthetic Perception, Mountain Retreat Lifestyle Experience, Emotional Healing and Spiritual Dwelling, and Symbolic Communication and Place Identity. The third section collected demographic and stay-related information. All measurement items were assessed using a five-point Likert scale, ranging from 1 = "strongly disagree"to 5 = "strongly agree." The formal questionnaire was distributed online through tourism-related WeChat groups, QQ groups, and Xiaohongshu private messages. Responses from individuals without Zhangjiajie homestay experience or with uncertain stay locations were excluded.

4.2. Demographic Profile of Respondents

A total of 208 questionnaires were collected from respondents with prior experience staying in Zhangjiajie homestays. After excluding three responses from participants whose most recent stay was more than two years ago, 205 valid questionnaires were retained to ensure relatively clear recall of landscape, spatial, and emotional experiences. This sample size was acceptable for validating a six-dimensional scale with 21 observed items. The response-to-item ratio was approximately 9.8:1, providing a sufficient basis for assessing reliability, convergent validity, and discriminant validity.

In terms of sample composition, 94 respondents were male, accounting for 45.85%, and 111 were female, accounting for 54.15%, indicating a relatively balanced gender distribution. The respondents were mainly concentrated in the 26-45 age group, with those aged 26-35 accounting for 35.12% and those aged 36-45 accounting for 31.22%, suggesting that the sample was dominated by young and middle-aged consumers. Regarding occupation, corporate employees accounted for the largest proportion, at 40.49%, followed by students, at 17.07%. In terms of education, respondents with bachelor's degrees and junior college degrees accounted for relatively high proportions, at 38.05% and 31.22%, respectively. Regarding travel motivation, vacation and leisure travel was the primary purpose of staying in Zhangjiajie homestays, accounting for 36.59%, followed by parent-child travel, accounting for 22.44%. Overall, the sample reflects the main characteristics of Zhangjiajie homestay consumers in terms of vacation, parent-child travel, leisure, and mountain retreat experiences.

4.3. Measurement Model Assessment

To evaluate the measurement quality of the landscape narrative scale for Zhangjiajie homestays, this study used SmartPLS to assess reliability, convergent validity, and discriminant validity. Since the six dimensions were derived from the previous CNN-informed visual-semantic analysis and grounded theory coding, the questionnaire analysis did not aim to re-explore the factor structure. Instead, it focused on examining whether each item could stably reflect its corresponding dimension.

The results of outer loadings showed that the standardized loadings of all 21 items on their corresponding dimensions were above 0.70, ranging from 0.789 to 0.870. These results indicate that all items adequately reflected their corresponding dimensions, with no low-loading items, suggesting good indicator reliability.

Table 4. Results of Outer Loadings

Dimensions	Items	Outer loadings
Atmospheric–Lightscape Perception	AL01	0.859
	AL02	0.821
	AL03	0.831
	AL04	0.854
Emotional Healing and Spiritual Dwelling	EH01	0.866
	EH02	0.869
	EH03	0.870
Homestay Spatial Aesthetic Perception	HS01	0.789
	HS02	0.826
	HS03	0.816
Mountain Retreat Lifestyle Experience	MR01	0.813
	MR02	0.793
	MR03	0.816
	MR04	0.830
Mountain-Water Ecological Landscape Perception	ME01	0.851
	ME02	0.846
	ME03	0.851
	ME04	0.853
Symbolic Communication and Place Identity	SC01	0.817
	SC02	0.851
	SC03	0.819

This study used Cronbach's α, rho_A, and composite reliability to examine the internal consistency of each dimension. The results showed that Cronbach's α ranged from 0.739 to 0.872, rho_A ranged from 0.740 to 0.872, and composite reliability ranged from 0.852 to 0.913, all exceeding the recommended threshold of 0.70. This indicates that the six dimensions had good internal consistency and measurement stability. In terms of average variance extracted, the AVE values ranged from 0.657 to 0.754, all above the recommended threshold of 0.50, indicating that the items converged well on their corresponding dimensions. Overall, the landscape narrative scale for Zhangjiajie homestays reached an acceptable level in terms of reliability and convergent validity.

Table 5. Reliability and Convergent Validity

	Cronbach's alpha	rho_a	rho_c	AVE
Atmospheric–Lightscape Perception	0.862	0.863	0.906	0.708
Emotional Healing and Spiritual Dwelling	0.837	0.839	0.902	0.754
Homestay Spatial Aesthetic Perception	0.739	0.740	0.852	0.657
Mountain Retreat Lifestyle Experience	0.829	0.831	0.886	0.661
Mountain-Water Ecological Landscape Perception	0.872	0.872	0.913	0.723
Symbolic Communication and Place Identity	0.773	0.773	0.868	0.687

To assess discriminant validity, this study used HTMT. All values were below 0.90, with a maximum of 0.846 between Homestay Spatial Aesthetic Perception and Mountain-Water Ecological Landscape Perception. Although the six dimensions are distinguishable, the relatively high values among spatial aesthetics, mountain-water ecology, lifestyle experience, and symbolic communication indicate conceptual proximity. This is theoretically reasonable because these dimensions jointly form the landscape narrative of Zhangjiajie homestays. They are therefore interpreted as interrelated first-order sub-dimensions of a higher-order landscape narrative construct.

Table 6. Discriminant Validity

Dimensions	AL	ES	HS	MR	ME
Emotional Healing and Spiritual Dwelling(ES)	0.743				
Homestay Spatial Aesthetic Perception(HS)	0.698	0.788			
Mountain Retreat Lifestyle Experience(MR)	0.710	0.737	0.749		
Mountain-Water Ecological Landscape Perception(ME)	0.751	0.717	0.846	0.764	
Symbolic Communication and Place Identity(SC)	0.743	0.788	0.780	0.814	0.806

5. Conclusion

5.1. Discussion

This study constructed and validated the structural dimensions of the landscape narrative of Zhangjiajie homestays by integrating Xiaohongshu UGC, CNN-informed hybrid visual-semantic analysis, grounded theory coding, and consumer questionnaire validation. The findings show that the landscape narrative of Zhangjiajie homestays is not a direct representation of natural scenery, but a structured system through which ecological landscapes, atmospheric lightscape, spatial aesthetics, lifestyle scenes, emotional meanings, and symbolic communication are organized into an accommodation-based narrative.

This finding is consistent with the view that tourist landscapes are not static physical objects, but meaning systems produced through ways of seeing, embodied experience, and spatial practice. In the context of Zhangjiajie homestays, mountains, water bodies, vegetation, mist, light, architecture, courtyards, terraces, and interiors are not isolated visual elements. Instead, they are reorganized through homestay spaces and transformed into livable and shareable landscape scenes. This extends landscape narrative research from destination-level landscapes to the micro-spatial context of accommodation.

The UGC-based coding results further reveal the visual and experiential complexity of Zhangjiajie homestay landscape narrative. Mountain-view integration, mountain-water enclosure, green coverage, and waterscape vacation indicate the ecological embeddedness of homestays, while mist imagery, seasonal atmosphere, morning and dusk landscapes, and Chinese landscape-painting imagery reflect their aesthetic and poetic organization. This echoes previous research showing that mountain landscapes in tourist narratives often carry multiple meanings related to natural heritage, visual aesthetics, healing, and cultural discovery [4].

This study also advances accommodation-related landscape narrative research. Prior studies have shown that accommodation spaces can contribute to place-making through architecture, spatial atmosphere, service rituals, and tourist evaluation [12]. By focusing on homestays, this study shows that accommodation spaces can transform external mountain-water resources into internal narrative structures through spatial design,

lifestyle arrangement, and platform-based visual communication. The questionnaire results further support the stability of this structure: all item loadings exceeded 0.70, and the reliability, convergent validity, and discriminant validity indicators reached accepted thresholds. Thus, the proposed scale provides a measurable foundation for future research on homestay landscape narrative and tourism accommodation experience.

5.2. Research Conclusion

Based on 119 short videos and 2,231 images, this study extracted 21 measurement indicators of the landscape narrative of Zhangjiajie homestays and grouped them into six dimensions: Mountain-Water Ecological Landscape Perception, Atmospheric-Lightscape Perception, Homestay Spatial Aesthetic Perception, Mountain Retreat Lifestyle Experience, Emotional Healing and Spiritual Dwelling, and Symbolic Communication and Place Identity. The SmartPLS results based on 205 valid questionnaires showed satisfactory measurement quality, with outer loadings ranging from 0.789 to 0.870, Cronbach's α from 0.739 to 0.872, composite reliability from 0.852 to 0.913, AVE from 0.657 to 0.754, and a maximum HTMT value of 0.846. Theoretically, this study contributes by extending landscape narrative research to the homestay context, developing an operational measurement framework, and proposing a research pathway of "visual identification-indicator extraction-dimension aggregation-measurement validation."

5.3. Implication

This study offers three practical implications for Zhangjiajie homestay development and destination communication. First, homestay operators should shift from facility-oriented design to landscape narrative scene creation. Spatial elements such as viewing windows, terraces, courtyards, walking paths, waterscapes, and lighting can strengthen the integration between homestays and Zhangjiajie's mountain-water environment, while mist, sunrise, post-rain scenery, and seasonal changes can be transformed into recognizable aesthetic scenes.

Second, homestay products should connect spatial aesthetics with Mountain Retreat Lifestyle Experience. Rather than emphasizing accommodation quality alone, operators should design lifestyle-oriented experiences such as courtyard leisure, tea brewing, mountain walking, parent–child nature activities, local cuisine, and Tujia cultural experiences. These activities can enhance the experiential and emotional value of homestays.

Third, Zhangjiajie homestays should strengthen symbolic communication in social media contexts. Brand signs, check-in scenes, local cultural elements, and seasonal landscape content can improve visual distinctiveness and transform individual sharing into destination image communication. Destination managers may also incorporate boutique homestays into Zhangjiajie's broader tourism communication system, guiding them to develop differentiated landscape expressions and stronger local cultural embeddedness.

5.4. Limitation

This study has several limitations. First, the UGC data were collected mainly from Xiaohongshu, and future research may compare other platforms such as Tik Tok, Ctrip.

Second, the analysis focused primarily on visual content, while subtitles, comments, background music, and interaction data were not fully incorporated. Third, the questionnaire sample contained 205 valid responses and was mainly used for measurement validation; future studies may expand the sample size and conduct cross-destination comparisons. Finally, this study validated the dimensional structure of the landscape narrative scale, but did not examine causal paths among dimensions or their effects on revisit intention, recommendation intention, or sharing intention. Future research may use structural equation modeling to further examine these relationships.

References

[1] UNESCO World Heritage Centre. Wulingyuan scenic and historic interest area [Internet]. Paris: UNESCO World Heritage Centre; [cited 2026 May 15]. Available from: https://whc. unesco.org/ en/list/640/

[2] Hunan Provincial People's Government. Zhangjiajie city: act as the "lead goose" for the Greater Zhangjiajie International Tourism Area [Internet]. Changsha: Hunan Provincial People's Government; 2026 Jan 27 [cited 2026 May 15]. Available from: https://www.hunan.gov.cn/hnszf/hnyw/szdt/202601/t20260127_33902825.html.

[3] Georgescu Paquin A, Cerdan Schwitzguébel A. Analysis of Barcelona's tourist landscape as projected in tourism promotional videos. International Journal of Tourism Cities. 2021;7(2):257-277.

[4] Mele E, Egberts L. Exploring travel blogs on tourism and landscape heritage: representations of the Swiss Alps. Journal of Heritage Tourism. 2023;18(6):785-806. doi:10.1080/1743873X.2023.2237617

[5] Yuan X, Yang R, Zhu C, Xue K. Tourists' narrative engagement and the multidimensional construction of historic urban landscapes: exploring the role of narration in place attachment. Landscape Research. 2024;49(6):773-786. doi:10.1080/01426397.2024.2349543

[6] Rakhmanova L. Intertwined narratives: nature tourism in the context of forced settlers' history in Western Siberia. Landscape Research. 2024;49(7):986-1000. doi:10.1080/01426397.2024.2383471

[7] Strauser J, Stewart WP, Evans NM, Stamberger L, van Riper CJ. Heritage narratives for landscapes on the rural-urban fringe in the Midwestern United States. Journal of Environmental Planning and Management. 2019;62(7):1269-1286. doi:10.1080/09640568.2018.1492908

[8] Schaal-Lagodzinski T, König B, Riechers M, Heitepriem N, Leventon J. Exploring cultural landscape narratives to understand challenges for collaboration and their implications for governance. Ecosystems and People. 2024;20(1):2320886. doi:10.1080/26395916.2024.2320886

[9] González-Álvarez D. Rethinking tourism narratives on the cultural landscapes of Asturias, Northern Spain, from the perspective of landscape archaeology: do archaeologists have anything to say? Landscape Research. 2019;44(2):117-133. doi:10.1080/01426397.2017.1413174

[10] Aschenbrand E, Cisani M, Meneghello S, Chance J. Heritage and landscape narratives: how rural stereotypes are shaping UNESCO biosphere reserves. Tour Geogr. 2025 Oct 29:1-18.

[11] Arcos-Pumarola J, Osácar Marzal E, Llonch-Molina N. Literary urban landscape in a sustainable tourism context. Hum Geogr. 2018;12(2):175-89.

[12] Thirumaran K, Mohammadi Z, Azzali S, Eijdenberg EL, Donough-Tan G. Transformed landscapes, tourist sentiments: the place making narrative of a luxury heritage hotel in Singapore. J Herit Tour. 2023;18(2):243-64.

[13] Zhao X, Wang L, Zhang Y, Han X, Deveci M, Parmar M. A review of convolutional neural networks in computer vision. Artif Intell Rev. 2024;57(4):99.

[14] Zou Z, Chen K, Shi Z, Guo Y, Ye J. Object detection in 20 years: a survey. Proc IEEE. 2023;111(3):257-76.

[15] Ali ML, Zhang Z. The YOLO framework: a comprehensive review of evolution, applications, and benchmarks in object detection. Computers. 2024;13(12):336.

[16] Edozie E, Shuaibu AN, John UK, Sadiq BO. Comprehensive review of recent developments in visual object detection based on deep learning. Artif Intell Rev. 2025;58(9):277.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)

© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261080

The Application of AI and Innovation Efficiency in Small and Medium-Sized Enterprises: A Perspective Based on Dynamic Capabilities and Resource Reallocation

Yichen Wu[a,1]
[a] *Shanghai Maritime University, China*

Abstract. Using Chinese listed SME data (2019–2024), we find that AI application significantly enhances innovation efficiency even under resource constraints. Dynamic capabilities play an inverted U-shaped mediating role: their mediation rises then falls after a "matching threshold". External market competition exerts a U-shaped moderation—only intense competition forces effective AI use.

Keywords. AI application, SME, Innovation Efficiency, Dynamic Capabilities

1. Introduction

Artificial Intelligence (AI) is the core driving force behind the Fourth Industrial Revolution, and digitalization, intelligentization, and networking have become the primary directions of development in this era. Amid the Industry 4.0 revolution, the integration of high-tech innovations and industrial sectors is growing increasingly close. Enterprises are accelerating the adoption and promotion of AI technologies to boost productivity, optimize operational models, and enhance their capacity for innovation and competitiveness. Since innovation efficiency serves as a key indicator of this capability, exploring the mechanisms and moderating factors through which AI applications influence corporate innovation efficiency holds significant practical importance.

Research on how AI influences firm-level innovation efficiency remains scarce, particularly for small and medium-sized enterprises (SMEs). Most existing work continues to weigh the overall benefits and drawbacks of AI from a macro lens—for instance, financial development [1] and regional environment [2] have been linked to innovation efficiency. At the micro level, factors such as equity structure [3] and managerial characteristics [4] are also recognized as determinants of corporate innovation. Yet, empirical studies that specifically examine AI as a technological factor are still limited. Drawing on dynamic capabilities theory, this study empirically tests its mediating role between AI adoption and SME innovation efficiency, aiming to identify the underlying channels at play.

[1] Corresponding Author: Yichen Wu, wuyichen020501@163.com.

2. Theoretical background and hypothesis development

2.1. AI and Innovation Efficiency in SME

Innovation efficiency reflects how well a firm translates a given set of innovation inputs into outputs[5]and is regarded as a core measure of innovation capacity. Prior studies on AI and corporate innovation efficiency largely indicate that AI application contributes to higher efficiency[6][7], with more intensive AI adoption coinciding with stronger innovation efficiency. Accordingly, we propose the following hypothesis: H1: AI application exerts a positive effect on the innovation efficiency of SMEs.

2.2. AI and the Dynamic Capabilities of SME

Teece (2007)[8]conceptualizes dynamic capabilities as a firm's capacity to coordinate, integrate, and reconfigure internal and external resources in response to rapidly shifting environments, thereby securing sustainable competitive advantage. This construct explains how firms sustain competitiveness under high environmental volatility. Organizations with strong dynamic capabilities are better positioned to spot relevant AI applications, embed them into existing workflows, and continuously recalibrate their deployment to promote ongoing innovation[9].

Accordingly, we advance Hypothesis H2: AI application enhances the dynamic capabilities of SMEs.

As for the underlying mechanism, WANG S (2024)[10] finds that AI adoption reshapes innovation efficiency by transforming human capital, innovation models, and the market environment. Introducing dynamic capabilities into this framework, we further propose Hypothesis H3: AI affects SME innovation efficiency by strengthening dynamic capabilities.

2.3. The regulatory role of resource reallocation

Resources are a critical factor in the survival and development of enterprises; the conduct of any business activity depends on the effective allocation of resources. Innovation is inherently resource-intensive, typically requiring substantial R&D funding, highly skilled talent, and advanced technological equipment as a foundation. However, market resources are finite, and market competition influences the reallocation of these resources. Therefore, we propose Hypothesis H4: Market competition moderates the relationship between resources and innovation efficiency in SMEs.

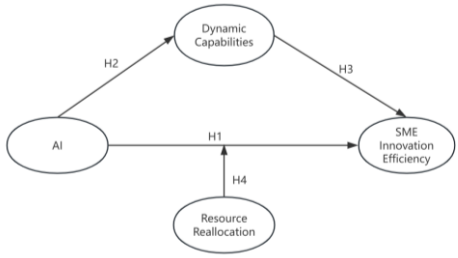


Figure 1. Research design.

3. Data and methodology

3.1. Econometric mode

This paper conducts an empirical study using data on listed small and medium-sized enterprises in China from 2019 to 2024, sourced from the CSMAR database. Based on the above analysis, the basic model is constructed as follows:

$$IE_{it} = \alpha + \beta AI_{it} + \gamma Control_{it} + \mu_{it} + \varepsilon_{it} \quad (1)$$

IE_{it} represents the innovation efficiency of listed small and medium-sized enterprises (SMEs), AI_{it} is the core explanatory variable, and $Control_{it}$ is the control variable considered in this paper; μ_{it} denotes the two-way fixed effects for both firm and time; ε represents the random error term; and the coefficient β indicates the direction and magnitude of the impact of AI adoption levels on the innovation efficiency of SMEs.

3.2. Variables description

The dependent variable (IE) is measured by the ratio of innovation inputs to innovation outputs. We calculate corporate innovation efficiency using the formula "ln(number of invention patents published in the current year + 1) / ln(R&D expenditure in the current year + 1)." The application of AI as an explanatory variable is measured by creating a dictionary of AI-related terms, scraping companies' annual reports, and calculating the frequency of relevant terms. This is expressed as (total frequency of AI-related terms / total number of terms in annual reports) * 100. For the mediating variable of dynamic capabilities, we construct an index to measure dynamic capabilities across three dimensions: absorption, adaptation, and innovation. This study uses the Herfindahl-Hirschman Index (HHI) to calculate the degree of market competition, representing resource reallocation. This study selects equity concentration, profitability, financial leverage, and ownership structure as control variables.

4. Results and discussion

4.1. Baseline Results

All empirical analyses were conducted using Stata. Table 1 reports the descriptive statistics of the variables. To evaluate the effect of AI adoption on SME innovation efficiency, we estimated the baseline model specified in Equation (1); the results are presented in Table 2. Column (1) shows that AI application has a significant positive effect on innovation efficiency. This finding remains robust in Column (2) after controlling for additional covariates, thus lending empirical support to Hypothesis 1.

Table 1. DESCRIPTIVE STATISTICS

Variable	Obs	Mean	Std.dev.	Min	Max
IE	7152	.377	.133	0	.637
AI	7152	.02	.043	0	.489
DC	7152	-.0006	.605	-.7444	2.440
Lev	7152	.388	.188	.014	1.797
Roa	7152	3.806	11.518	-62.76	419.41
Shrcr	7152	.284	.13	.011	.9
Age	7152	3.309	.246	1.946	4.06
Soe	7152	.157	.364	0	1
MC	7152	.103	.141	.015	1

Table 2. BASELINE RESULTS

	(1) IE	(2) IE
AI	0.177*	0.183*
	(2.46)	(2.54)
Lev		0.00191
		(0.78)
ROA		-0.000102
		(-067)
Shrcr		-0.0227
		(-0.83)
age		0.03369
		(0.85)
_cons	0.368***	0.261*
	(148.97)	(2.06)
Frim FE	Yes	Yes
Time FE	Yes	Yes
Obs	7152	7152
R^2	0.0607	0.0610

4.2. Robustness Tests

To address potential reverse causality and endogeneity from omitted variables, we instrumented AI adoption with its one- and two-period lagged values and estimated a dynamic panel model via system GMM, which also uses lagged dependent variables as instruments. The results remained consistent with our predictions. For a robustness check, we reconstructed the innovation efficiency measure (te) using SFA after recalculating innovation output following the patent weighting scheme of Bereskin et al. (2016)[11]and Tong et al. (2014)[12](invention, utility model, and design patents

weighted at 1, 1/2, and 1/3, respectively). Re-estimating the model with this alternative efficiency measure yielded results that continued to support the hypotheses.

4.3. Mediating effect of DC

To examine the mediating role of dynamic ability, this study established a mediation model and conducted a regression analysis. The results of Table 3 indicate that AI has a significant positive moderating effect on dynamic ability, consistent with Hypothesis H2. Regarding Hypothesis H3, we first conducted stepwise regression to test for mediation. The results indicated that DC exhibits a full mediation effect; however, this was not supported by the Bootstrap test, as the confidence interval for the mediation effect included zero, suggesting that the linear mediation effect was not robust. Therefore, we further examined whether a nonlinear relationship existed. To test whether dynamic capabilities (DC) mediate the relationship between AI and innovation efficiency in SMEs in a nonlinear manner, we constructed a structural equation model that included a squared term for AI and conducted nonlinear tests after centering the AI application level (AI_c). The results indicate a significant inverted U-shaped relationship between AI adoption levels and dynamic capabilities (DC) ($a_1 = 8.69$, $p < 0.001$; $a_2 = -27.47$, $p < 0.001$), which further supplements and refines Hypothesis H2.

Table 3. Nonlinear testing

	Coefficient	std. err.	z	P>|z|	[95% conf. interval]	
Structural						
DC						
AI_c	8.694241	.3700206	23.50	0.000	7.969014	9.419467
AI_c2	-27.47112	2.71517	-10.12	0.000	-32.79276	-22.14949
_cons	.0385446	.007724	4.99	0.000	.0234058	.0536834
IE						
DC	-.0188843	.0027464	-6.88	0.000	-.0242672	-.0135014
AI_c	.2495226	.0439849	5.67	0.000	.1633138	.3357314
_cons	.3769639	.0015634	241.11	0.000	.3738996	.3800281
var(e.DC)	.3194825	.0053425			.309181	.3301272
var(e.IE)	.017482	.0002923			.0169183	.0180644

4.4. The Regulatory Role of Market Competition

When testing for moderation effects, the variables were centered and lagged by one period. The moderation effect of HHI was not significant in the one-way interaction test. To test for nonlinear effects, as shown in Table 4 (1), the results remain insignificant. However, upon segmenting the industry and focusing the analysis solely on small and medium-sized enterprises (SMEs) in the manufacturing sector, as shown in Table 4(2), both the first-order and second-order interaction effects are significant. Further graphical analysis of the nonlinear relationship, as depicted in Figure 2, indicates that market competition exerts a positive U-shaped moderating effect on the process by which AI applications influence the innovation efficiency of SMEs in the manufacturing sector.

One possible explanation for this is that AI, as a general technology that requires continuous investment, system integration, and organizational support, is a resource with high entry barriers for small and medium-sized enterprises. Without sufficient pressure

to drive change, these enterprises tend to maintain the status quo, and their AI investment remains at the pilot or shallow application stage. When a certain threshold is reached, the resource restructuring and dynamic capability building capabilities of small and medium-sized enterprises are activated, enabling them to cross the "resource constraint" threshold and transform potential technological benefits into actual innovation efficiency.[2]

Table 4. Regulatory function

	(1) LIE	(2) LIE
Lc_AI	0.120	-0.322*
	(0.88)	(-2.03)
Lc_HHI	0.0256	0.0424
	(0.60)	(0.77)
c.Lc_AI#c.Lc_HHI	-0.00881	-3.455*
	(-0.01)	(-2.22)
c.Lc_AI#c.Lc_HHI_sq	5.881	14.37*
	(1.44)	(2.16)
ROA	-0.000391	-0.000350
	(-1.91)	(-1.40)
Lev	0.0292	0.0252
	(1.48)	(1.14)
Shrcr	0.00684	-0.00484
	(0.17)	(-0.10)
age	0.00303	0.00917
	(0.04)	(0.11)
Time FE	Yes	Yes
Frim FE	Yes	Yes
_cons	0.355	0.362
	(1.59)	(1.50)
Obs	5680	4343
R^2	0.0152	0.0122

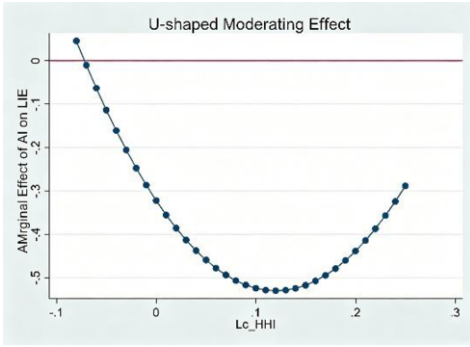


Figure 2. Graph of Nonlinear Control Effects.

[2] AR(1) p=0.000,AR(2) p=0.476, Hansen J p=0.163

5. Conclusion

Based on empirical data from listed small and medium-sized enterprises (SMEs) in China from 2019 to 2024, this paper examines the mechanisms through which AI applications influence the innovation efficiency of SMEs. It focuses on analyzing the mediating role of dynamic capabilities and the moderating role of resource reallocation. The key findings are summarized below. First, AI adoption significantly drives SME innovation efficiency. Baseline regression results confirm that AI technology markedly improves efficiency; even resource-constrained SMEs can realize meaningful innovation leaps through AI empowerment. Second, dynamic capabilities mediate this relationship in an inverted U-shaped manner. As AI application intensifies, the mediation effect first rises and then declines. This pattern points to a "matching tipping point" between technological empowerment and organizational capabilities: when AI adoption passes a critical threshold, existing dynamic capabilities may fail to co-evolve, creating fresh bottlenecks. Such a finding extends nonlinear interpretations of dynamic capabilities theory in the context of digital transformation. Third, external market competition exerts a positive U-shaped moderating effect on the AI–innovation efficiency link among manufacturing SMEs. Under low competitive pressure, firms lack the urgency to exploit AI for innovation; once competition intensity exceeds a critical level, external pressure compels firms to deploy AI resources more effectively, turning the moderation positive. This article has certain limitations. For instance, the measurement methods of the independent and dependent variables may be subject to deviations due to factors such as time.

References

[1] Hsu PH, Tian X, Xu Y. Financial development and innovation: cross-country evidence. Journal of Financial Economics. 2014;112(1):116-135.
[2] Komikado H, Morikawa S, Bhatt A, Kato H. High-speed rail, inter-regional accessibility, and regional innovation: evidence from Japan. Technological Forecasting and Social Change. 2021;167:120697.
[3] Wang Y, Zhao J. Hedge funds and corporate innovation. Financial Management. 2015;44(2):353-385.
[4] Sunder J, Sunder SV, Zhang J. Pilot CEOs and corporate innovation. Journal of Financial Economics. 2017;123(1):209-224.
[5] Cruz-Cázares C, Bayona-Sáez C, García-Marco T. You can't manage right what you can't measure well: technological innovation efficiency. Research Policy. 2013;42(6-7):1239-1250.
[6] Jiang L, Xuan Y, Zhang K. Unlocking innovation potential: the impact of artificial intelligence transformation on enterprise innovation capacity. European Journal of Innovation Management. 2024 Nov 26. doi:10.1108/EJIM-07-2024-0809
[7] Li C, Xu Y, Zheng H, Wang Z, Han H, Zeng L. Artificial intelligence, resource reallocation, and corporate innovation efficiency: evidence from China's listed companies. Resources Policy. 2023 Mar;81:103324. doi:10.1016/j.resourpol.2023.103324
[8] Teece DJ. Explicating dynamic capabilities: the nature and microfoundations of (sustainable) enterprise performance. Strategic Management Journal. 2007;28(13):1319-1350.
[9] Warner KS, Wäger M. Building dynamic capabilities for digital transformation: an ongoing process of strategic renewal. Long Range Planning. 2019;52(3):326-349.
[10] Wang S, Huang X, Xia M, Shi X. Does artificial intelligence promote firms' innovation efficiency: evidence from the robot application. Journal of the Knowledge Economy. 2024 Jan 25;15(4):16373-16394. doi:10.1007/s13132-023-01707-w
[11] Bereskin FL, Campbell TL, Hsu PH. Corporate philanthropy, research networks, and collaborative innovation. Financial Management. 2016;45(1):175-206.
[12] Tong TW, He W, He ZL, Lu J. Patent regime shift and firm innovation: evidence from the second amendment to China's patent law. In: Academy of Management Proceedings. 2014;2014(1):14174.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261081

Research on the Influencing Mechanism of Participation Intention in Immersive Red Tourism: A Case Study of the Guangdong–Hong Kong–Macao Greater Bay Area

Xiaoqi Huang, Hao Wu, Shuang Li[1], Weiqin Wu, Jingqi Lei
Guangdong Polytechnic of Industry and Commerce, Guangzhou, China

Abstract: Red tourism has become a key driver of cultural inheritance and the integration of culture and tourism. As an innovative form of red tourism, immersive red tourism serves as an effective means of deepening the public's understanding of and emotional connection to revolutionary culture. Based on a survey of 1,064 tourists in the Guangdong–Hong Kong–Macao Greater Bay Area (GBA), this study investigates tourists' participation intention in immersive red tourism. The findings reveal that the red tourism market in the GBA is characterized predominantly by low expenditure, with a smaller but notable high-end segment. In addition, problems related to insufficient resource integration and lagging technological application remain prominent. Social influence and Participation Motivation significantly and positively affect tourists' participation intention, whereas technology acceptance shows a negative effect, largely due to poor user experience and a misalignment between technological offerings and tourists' actual needs. Based on these findings, this study proposes strengthening regional collaboration and joint development to enhance the overall influence of immersive red tourism, while optimising the application of immersive technologies to promote high-quality development and practical innovation. The study highlights an interdisciplinary analytical perspective and a focus on regional characteristics, expands the understanding of tourists' participation in red tourism, and provides practical insights for the integrated development of culture and tourism in the GBA.

Keywords: Immersive red tourism; Tourists' participation intention; Embodied cognition theory

1. Introduction

From the "Bay Area Development Plan" proposed in the Planning Study on the Coordinated Development of the Greater Pearl River Delta Townships in 2009, to the comprehensive launch of regional synergy in the Outline of the Plan for the Development of the Guangdong—Hong Kong—Macao Greater Bay Area in 2019, the GBA has become a core vehicle for the country to deepen reforms and promote regional coordination. The GBA consists of nine municipalities—Guangzhou, Shenzhen, Zhuhai, Foshan, Huizhou, Dongguan, Zhongshan, Jiangmen, and Zhaoqing—plus the two Special Administrative Regions of Hong Kong and Macao, forming a "9+2" city cluster.

[1] Corresponding Author: Shuang Li, 466125908@qq.com.

The Guangdong–Hong Kong–Macao Greater Bay Area Culture and Tourism Development Plan highlights red tourism as an important sector, committing to building a distinctive red tourism brand in the GBA and to promoting regional cultural exchanges and economic cooperation. These mutually reinforcing policies provide strong policy support for integrating revolutionary cultural resources in the GBA, innovating red tourism development models, and achieving deep integration between the regional economy and the red tourism industry, thus opening up new opportunities for industrial development. As a strategic hub of China's "dual circulation" strategy, the GBA has seen increasing demand for upgraded consumption, which is closely aligned with its regional development objectives. In particular, urban residents' growing desire for meaningful cultural and spiritual experiences has created market opportunities for the development of immersive red tourism, thereby meeting residents' cultural and psychological needs.

According to preliminary estimates by the Guangdong Provincial Department of Culture and Tourism, the 13 classic red tourism scenic spots under monitoring received 1.653 million visitors during the eight-day Spring Festival holiday in 2025 (January 28 to February 4, from Lunar New Year's Eve to the seventh day of the new year), a year-on-year increase of 5.4%. Although this data shows that red tourism in the GBA has gained a certain degree of popularity, as stated in the Proposal on Promoting the High-Quality Development of Red Tourism in Our Province, the current development of red tourism in the GBA is still constrained by challenges. These challenges are mainly reflected in the limited red tourism formats and homogeneous red tourism offerings, which leaves consumption potential untapped; the insufficient output of red tourism and underutilization of united front advantages; and the inadequate development of united front advantages; insufficient development and utilisation of revolutionary cultural resources; and the urgent need to upgrade propaganda and communication methods [1]. Source: Proposal on Promoting the High-Quality Development of Red Tourism in Our Province, Guangdong Provincial Political Consultative Conference website. This study systematically analyses the intention of tourists from Guangdong, Hong Kong, and Macao to participate in immersive red tourism and its influencing factors from a micro-level empirical perspective, with a view to enhancing the appeal of red tourism tours, deepening tourists' understanding and emotional resonance of red culture, and helping upgrade the red tourism industry.

2. Theoretical framework and research hypotheses

2.1. Theoretical framework

Embodied cognition theory, a pivotal paradigm in contemporary cognitive science, addresses the limitations of traditional disembodied cognition, which separates the body, environment, and cognition. This theory posits that individual cognition is not an isolated, static process of brain information processing but rather a holistic, dynamically evolving construction of meaning rooted in external environments. The essence of cognition arises from interactions between the body and its environment, where sensory perception and behavioral engagement profoundly influence cognitive processing, emotional generation, and value formation. Embodied cognition emphasizes the intrinsic unity of cognition, emotion, motivation, and behavior, effectively fostering coordinated development of cognitive and emotional functions while enhancing personal agency and behavioral competence. In immersive red tourism contexts, the embodied cognition mechanism can

be broken down into four progressive dimensions: participation motivation, acceptance, embodied experience, and social influence. Participation motivation stimulates individuals' intrinsic drive to participate; acceptance fosters positive attitudes toward red culture; embodied experience facilitates the formation of core cultural meanings through immersive bodily engagement; social influence further shapes perceptions within group contexts. These multidimensional synergies collectively drive the formation of tourists' participation intention in immersive red tourism experiences (Wilson, 2002) [2].

From an international scholarly perspective, Chinese immersive red tourism represents a unique immersive memory experience paradigm of Eastern revolutionary heritage tourism. It synthesizes general sensory and technological immersive mechanisms and cultural heritage narrative immersion mechanisms [3]. Distinct from conventional entertainment-centric immersive tourism, its primary purpose centers on reconstructing revolutionary collective memory, fostering cultural identity, and strengthening national emotional attachment. This paradigm thus forms a localized cultural tourism model that conforms to universal immersive experience principles while retaining China's cultural distinctiveness [4].

2.2. Research hypotheses

(1) Participation Motivation and the Purpose of Tourist Participation in Immersive Red Tourism

In the field of consumption, Participation Motivation refers to the psychological drive or intrinsic need that leads tourists to actively engage in a certain consumption behaviour, activity, or interactive process. It reflects why tourists are willing to invest time, energy, or resources in a consumption-related experience rather than passively accepting a product or service. Pratisto et al. found in a systematic review that immersive technologies significantly increased tourists' emotional engagement and behavioural intentions. Xu Tao et al. pointed out that, on the one hand, it is the result of internal psychological drive, and on the other hand, it is the result of the attraction of the external world [5]. Zhao Weihong and Wang dong summarised the main motivations for tourists to participate in virtual brand communities as: searching for effective information, obtaining preferential treatment, seeking offline interactions, sharing post-purchase experiences and maintaining information quality [6]. The feedback from tourists in the research on the motivation for participation in immersive red tourism verified this view, and they also pointed out key elements such as patriotic education, based on which the following hypotheses were formulated:

M1: Gaining immersive red tourism experiences positively affects tourists' understanding of revolutionary history.

M2: Family and group participation in immersive red tourism serves as an effective form of patriotic education.

M3: Active participation in red tourism activities significantly strengthens tourists' sense of cultural identity.

(2) Technology Acceptance and Tourists' Acceptance of Immersive Red Tourism

In 1995, Williams and Hobson first explored VR's feasibility in tourism, proposing that it could overcome spatial and temporal limitations via scene simulation. This work laid the foundation for subsequent research on immersive tourism and provided the basis for perceived usefulness. Weber—Sabil and Han proposed the “XR Technology Layering Model,” which emphasises the need to hierarchically match technology with tourists’

needs and provides methodological support for the precise design of immersive tourism in terms of perceived ease of use. From the perspective of tourists, Chen Guofeng and Feng Jun pointed out that red tourism in the GBA suffers from scattered resources and homogenised experiences. These problems need to be addressed through cross-regional linkage and technological empowerment, which in turn requires tourists to be willing to use technology. In immersive red tourism, tourists' acceptance of technology is determined by three factors: ease of use, realism of scene reproduction, and attitude toward the technology. Based on the above, we propose the following hypotheses:
T1: The adoption of VR/AR technology positively influences tourists' perceived usefulness by vividly restoring historical scenes.
T2: User-friendly immersive devices enhance tourists' perceived ease of use, leading to a smoother experience.
T3: The launch of new immersive technology projects at scenic spots positively influences tourists' attitudes toward technology and their willingness to participate.

(3) Embodied Experience and Tourist Participation in Immersive Red Tourism Experiences
Embodied experience refers to the holistic experience gained by individuals through body perception, movement participation, and environmental interaction, emphasising the inseparability of cognition, emotion and behaviour from the body. In the field of immersive tourism research, Wang Jinye and Feng Yuxiao used the rooted analysis method to analyse the logic of 'scenario narrative' in Jiangxi Wuyuan immersive rural tourism, and proposed that embodied experience needs to take into account both sensory stimulation and cultural identity. Cai Jialing, Xie Linyuan, and other scholars analysed the development status quo, path and communication strategy of immersive experiences in red tourism scenic spots in Guangxi, and put forward measures to enhance the immersive experience of red tourism scenic spots by innovating the form of experience and strengthening publicity and promotion. Xie Junlin, Huang Youmei and other scholars explored the development strategy of Guangxi red tourism immersive experience based on the theory of heart flow, emphasising the optimisation of the experience process and the increase of interactive links. Based on the above discussion, the following hypotheses are proposed:
E1: Physical interactions, such as simulated battles, enhance tourists' realistic experience of revolutionary history.
E2: Multi-sensory stimulation (vision, hearing, touch, etc.) effectively enhances tourists' sense of immersion.
E3: Role-playing tasks foster emotional resonance between tourists and historical figures.

(4) Social Influences and Tourists' Participation Intention in Immersive Red Tourism
Social influence refers to the adaptive changes in cognition, emotion and behaviour of an individual or a group through the dynamic coupling of physical interaction, situational embeddedness and perceptual-motor systems, which are jointly shaped by the physical and social environments. Li—Ling et al. pointed out that tourists' perceptions and evaluations of tourism digital experiences can directly influence tourists' civic behaviours and indirectly influence the emergence of tourists' civic behaviours [7]. Li Ruyou et al. pointed out that tourists should be encouraged to become producers and disseminators of tourism information, and provide social support to each other in the process of participating in the co-creation of value [8]. The above research results, combined with tourists' research on the social impact of immersive red tourism, suggest that:

S1: Recommendations from friends and relatives positively influence tourists' likelihood of choosing immersive red tourism.
S2: Positive social media reviews (e.g., RedNote) significantly influence tourists' travel decisions.
S3: Sharing immersive red tourism experiences on social platforms (e.g., WeChat Moments) positively contributes to tourists' perceived social acceptance.

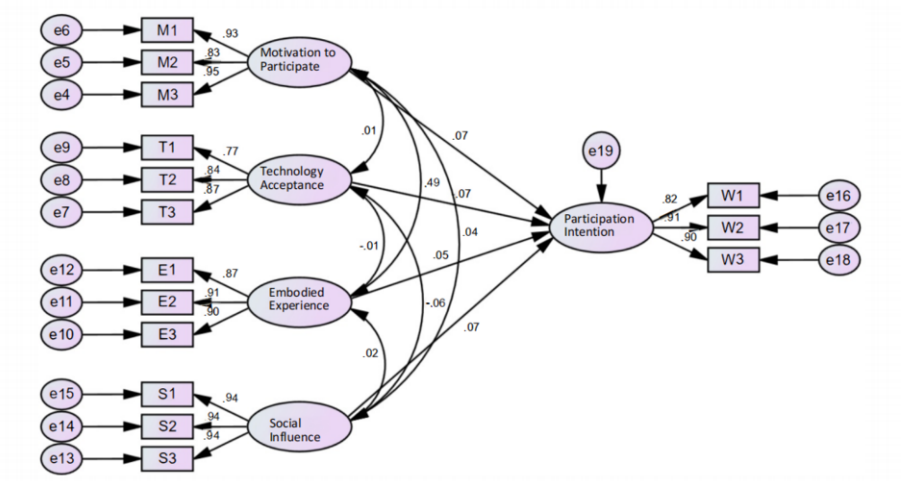


Figure 1. Conceptual model of factors influencing Participation Intention in immersive red tourism

3. Study Subjects and Sample Statistics

This study establishes a standardized research framework for immersive red tourism in the GBA, encompassing questionnaire development, scale adaptation, pilot surveys, sampling, and bias control. The questionnaire comprises three modules organized around the structure of "basic information—current development status—willingness exploration," covering demographic characteristics, red tourism consumption behaviors, and immersive participation intention, with core variables measured using Likert scales. Research items were developed by incorporating insights from the Technology Acceptance Model (TAM), immersive experience assessment scales, and established domestic studies on red tourism and Greater Bay Area tourism services, followed by rigorous translation and semantic calibration to align with the study context. Prior to formal implementation, pilot tests were conducted to refine item formulations and questionnaire structure, enhancing measurement validity. The study targeted potential tourists aged 18 and above who had traveled at least once in the past year, employing stratified, random, and quota sampling methods proportionally allocated based on population distribution across the 11 cities of the Greater Bay Area, supplemented by data collected through both online and offline channels. All procedures strictly adhered to academic ethics principles, ensuring anonymous responses and voluntary participation while safeguarding respondent privacy. Additionally, a common method bias assessment was implemented by scrambling question sequences and conducting multi-scenario sampling to mitigate source bias, thereby ensuring sample quality and data reliability.

To reduce the impact of regional differences and population characteristics on sample representativeness, this study adopted a dual-track distribution mechanism combining "geographic stratification and population quota." Through integrated online and offline channels, questionnaires were targeted at the urban agglomeration of the GBA. Electronic questionnaires were created using the Tencent Questionnaire platform and simultaneously distributed via social media platforms such as WeChat and Alipay (for Hong Kong and Macao users). Paper questionnaires were mainly distributed at 18 cross-border transportation hubs, including Guangzhou-Shenzhen-Hong Kong High-Speed Rail stations and Zhuhai Hengqin Port, as well as at major red tourism sites in Guangzhou and its surrounding areas (e.g., the Peasant Movement Institute, and the Memorial Hall of the Third Congress of the Communist Party of China). A total of 1,064 questionnaires were collected. After data cleaning, 1,025 valid questionnaires were retained, yielding an effective response rate of 96.33%.

3.1. Reliability Analysis

Table 1. Reliability statistics

Cronbach's alpha	Number of Terms
0.892	34

Reliability statistics indicate that the scale's Cronbach's Alpha coefficient is 0.892, demonstrating high internal consistency and effective measurement reliability. With 34 items, the scale exhibits a comprehensive structure covering multiple dimensions and subscales, making it suitable for complex research contexts. The high Alpha coefficient (above 0.8) further confirms the scale's stability, providing a solid foundation for subsequent data analysis.

3.2. Validity Analysis

Table 2. KMO and Bartlett's Test

KMO Sample Appropriateness Measure		0.925
Bartlett's Test of Sphericity	Approx. Chi-Square	24087.285
	df	351
	Sig.	0.000

The KMO and Bartlett's tests yielded a KMO value of 0.925, indicating that the data are highly suitable for factor analysis. The Bartlett's test revealed a significance level of 0.000 (with 351 degrees of freedom and an approximate chi-square value of 24,087.285), further confirming significant correlations among variables, rejecting the assumption of variable independence, and supporting the appropriateness of factor analysis. This suggests high structural validity of the data, enabling effective extraction of latent factors and providing a solid foundation for subsequent analyses.

4. Analysis of Empirical Results

4.1. Analysis of Participation Intention in immersive red tourism based on Structural Equation Modelling (SEM)

(1) Confirmatory Factor Analysis

Table 3. Model fit indices for confirmatory factor analysis

Commonly used indices	χ^2/df	RMSEA	GFI	IFI	TLI	CFI
Fit criteria	<3	<0.08	>0.9	>0.9	>0.9	>0.9
Values	1.273	0.016	0.987	0.998	0.998	0.998

Based on the confirmatory factor analysis model fit indices presented in Table 3, the model demonstrates a high degree of goodness-of-fit. The ratio of chi-square to degrees of freedom (X^2/df) is 1.273, which is well below the cutoff criterion of 3, indicating a good fit between the model and the data.

Table 4. Convergent validity analysis

Latent variable	Item	Standardised Factor Loading	CR	AVE
Participation Motivation	M1	0.929	0.930	0.816
	M2	0.828		
	M3	0.949		
Technology acceptance	T1	0.949	0.918	0.788
	T2	0.837		
	T3	0.874		
Embodied experience	E1	0.862	0.918	0.788
	E2	0.908		
	E3	0.893		
Social influence	S1	0.937	0.957	0.881
	S2	0.943		
	S3	0.935		
Participation Intention	W1	0.82	0.911	0.773
	W2	0.911		
	W3	0.904		

Based on the convergent validity analysis data presented in Table 4, all latent variables demonstrate a high degree of convergent validity. For Participation Motivation, the estimated values of the three items (M1, M2, and M3) are uniformly high at 0.929, 0.828, and 0.949, respectively, indicating that these items are highly correlated with the Participation Motivation latent construct. Its composite reliability (CR) value is 0.930, and the average variance extracted (AVE) value is 0.816, both of which meet the established criteria for convergent validity (CR > 0.7, AVE > 0.5). For technology acceptance, the estimated values of the three items (T1, T2, and T3) meet the convergent validity criteria. For embodied experience, the estimated values, along with the CR and

AVE values of its three items (E1, E2, and E3), indicate strong convergent validity for this construct as well. The three items (S1, S2, and S3) measuring social influence also demonstrate high convergent validity. Finally, the three items (W1, W2, and W3) for participation intention satisfy the criteria for convergent validity.

Table 5. Discriminant validity analysis

	Participation Motivation	Technology acceptance	Embodied experience	Social influence	Participation Intention
Participation Motivation	0.903				
Technology acceptance	0.008	0.888			
Embodied experience	0.451***	—0.006	0.888		
Social influence	0.047	—0.061	0.005	0.939	
Participation Intention	0.091**	—0.055*	0.05*	0.065*	0.879
AVE	0.816	0.788	0.788	0.881	0.773

Note: ***p < 0.001, **p < 0.01, *p < 0.05.

As shown in Table 5, the correlation coefficients among the latent variables are substantially lower than the square roots of their corresponding average variance extracted (AVE) values, indicating satisfactory discriminant validity. The inter-construct correlations are consistently lower than the AVE square roots, and their associated significance levels ($p < 0.001$, $p < 0.01$, $p < 0.05$) further corroborate the reliability of the discriminant validity. These findings demonstrate that the model performs well in distinguishing between different latent constructs, thereby effectively supporting subsequent structural equation modelling (SEM) analysis.

(2) Structural Equation Modelling (SEM)

Table 6. Structural equation model fit indices

Commonly used indices	χ^2/df	RMSEA	GFI	IFI	TLI	CFI
Fit criteria	<3	<0.08	>0.9	>0.9	>0.9	>0.9
Values	1.290	0.017	0.987	0.998	0.997	0.998

The fit indices reported in Table 6 suggest that the structural equation model provides a satisfactory fit to the observed data. Taken together, these indices indicate that the structural equation model fits the data exceptionally well, thereby effectively supporting subsequent path analysis and hypothesis testing.

Table 7. Significance testing of path coefficients

Path			Standardized coefficients	Unstandardized coefficient	S.E.	C.R.	P
Participation Intention	_____	Participation Motivation	0.074	0.050	0.026	1.906	0.057
Participation Intention	_____	Technology acceptance	—0.071	—0.060	0.029	—2.055	0.040
Participation Intention	_____	Embodied experience	0.050	0.046	0.036	1.266	0.205
Participation Intention	_____	Social influence	0.068	0.049	0.024	2.053	0.040

The path coefficients reported in Table 7 reveal that Participation Motivation, technology acceptance, embodied experience, and social influence exert varying effects on participation intention. Specifically, technology acceptance exerts a significant negative effect on Participation Intention. This confirms the Innovation Resistance Theory (IRT), which posits that barriers to technology adoption significantly foster innovation resistance and suppress participation intention. In 2024, G. Sowmya, Aruna Polisetty, et al., building upon the IRT framework, employed qualitative interviews and a PLS-SEM model to identify eight major obstacles to technology adoption in metaverse cultural tourism: three functional barriers (use barriers, value barriers, risk barriers); four psychological barriers (traditional experience barriers, image barriers, technological vulnerability barriers, ideological barriers); and individual factors contributing to passive innovation resistance. Their findings demonstrate that these obstacles lead to tourists' technological resistance, creating a negative feedback loop between superficial technology acceptance perceptions and actual participation intention in cultural tourism experiences, thereby explaining the counterproductive inhibitory effect of overly digitised immersive technologies in heritage tourism contexts [9]. Additionally, social influence has a significant positive effect on participation intention, indicating that supportive factors within the external social environment can effectively enhance an individual's participation intention. However, the effects of Participation Motivation ($p = 0.057$) and embodied experience ($p = 0.205$) on Participation Intention did not reach statistical significance, suggesting that these relationships may be moderated by other potential factors or influenced by limitations within the dataset. This result aligns with the practical characteristics of research on participation intention in red tourism: tourist participation intention is influenced by numerous variables not incorporated into this model, including destination accessibility, scenic area infrastructure, regional cultural tourism policies, peer decision-making, family factors, holiday constraints, cultural tourism marketing communication, and macro-environmental factors. None of these variables were included in the model, resulting in low overall explanatory power. Additionally, baseline demographic variables (gender, age, occupation, income) and previous red tourism consumption behavior variables showed no significant impact, further reducing overall model fit. Nevertheless, the model identified three significant predictive variables and confirmed the direct effects of three core latent variables on participation intention, demonstrating the model's theoretical relevance.

Regarding the magnitude of the path coefficients, the effects of technology acceptance and embodied experience on participation intention are relatively weak, with standardised coefficients of -0.071 and 0.050, respectively. In contrast, the effect of social influence is relatively stronger (standardised coefficient = 0.068). The critical ratios (C.R.) further confirm the significance of technology acceptance and social influence; at —2.055 and 2.053, respectively, both exceed the absolute critical threshold of 1.96, thereby satisfying the criteria for statistical significance. Overall, the findings highlight the critical role of social influence in enhancing participation intention, whereas the negative effect of technology acceptance suggests a need to optimise technological design to mitigate its inhibitory effect on participation intention. A summary of the aforementioned analytical results is presented in Table 8.

Table 8. Paths with significant path coefficients

Serial number	Path			Significance
1	Participation Intention	<———	Participation Motivation	No
2	Participation Intention	<———	Technology acceptance	Yes
3	Participation Intention	<———	Embodied experience	No
4	Participation Intention	<———	Social influence	Yes

5. Conclusions and Implications

This study constructs a conceptual model of the mechanisms influencing participation intention in immersive red tourism based on embodied cognition theory and its analytical framework, conducts factor validation using questionnaire data, and draws the following conclusions: The red tourism market in the Guangdong–Hong Kong–Macao Greater Bay Area exhibits distinctive characteristics; social influence significantly enhances tourists' participation intention, whereas technology acceptance exerts a negative impact. These findings provide important insights for improving participation intention in immersive red tourism within the region, and propose management recommendations and practical approaches from three dimensions: strengthening social influence, optimizing technology acceptance, and promoting regional coordinated development.

First, the positive social impact should be leveraged to establish a socially driven immersive red tourism participation mechanism. Empirical results demonstrate that social influence has a significant positive effect on tourists' intention to participate. Accordingly, a systematic social dissemination matrix for red cultural tourism should be developed, integrating mainstream social platforms such as Xiaohongshu and TikTok to plan and produce immersive red-themed experiential content targeting the GBA. This will encourage tourists to share check-in records, role-playing experiences, and other user-generated content, thereby stimulating potential participants' engagement through peer influence and social recognition mechanisms. Additionally, group-oriented red tourism products should be prioritized, designed for schools, enterprises, institutions, and families, including immersive scenario-based games and interactive educational tours. By leveraging the social amplification effects of organizational and community networks, the target market can be effectively expanded. Furthermore, a tiered social dissemination mechanism should be established, incorporating incentives for sharing check-ins and co-creation activities centered around red-themed narratives to enhance tourists' perception of social value during immersive experiences, ultimately fostering sustained participation and recommendation behaviors.

Secondly, it is essential to address the negative impacts of technology adoption by achieving dual optimization in both technological adaptation and the core values of red culture. Empirical analysis demonstrates that technology acceptance has a significantly negative effect on tourists' willingness to participate, highlighting a structural mismatch between current technological offerings and actual visitor needs. To address this, we should adopt lightweight, user-friendly technological solutions, streamline operation procedures for VR/AR devices, deploy dedicated support staff, effectively mitigate physical discomfort such as dizziness and system lag, and alleviate tourists' resistance to technology use. In content design, emphasis should shift from purely entertainment-

oriented technologies to strengthening narratives rooted in red culture, positioning digital technologies as auxiliary tools for historical storytelling while striking a balance between technological presentation and historical authenticity—thereby resolving the fundamental misalignment between technological offerings and tourists' cultural expectations. Additionally, a tiered product supply system should be established to cater to the distinct technological needs across age groups, offering hands-free immersive experiences tailored for users with low tech acceptance levels. This approach will reduce technological barriers and expand the reach of immersive red tourism initiatives.

Third, leveraging the regional characteristics of the Guangdong–Hong Kong–Macao Greater Bay Area, we should promote cross-regional collaboration and differentiated development. To address current challenges such as the lagging integration of red tourism resources and homogenized technological applications, it is essential to fully utilize the spatial structure and institutional advantages of the "9+2" city cluster to establish a framework for coordinated cross-regional immersive red cultural tourism development. Core cities like Guangzhou and Shenzhen should take the lead in creating immersive inter-city red cultural tourism routes, standardize digital services—including multilingual tour guide systems and online booking platforms—to enhance overall service quality. Concurrently, collaboration with Hong Kong and Macao should be strengthened to develop cross-border red study-tour products, leveraging international social networks to amplify positive societal impact and inject sustained momentum into the high-quality development of red cultural tourism in the region.

Funding

Ministry of Education, Teachers of Ideological and Political Theory Classes in Colleges and Universities, Research Special Project on ‘Research on the Four-Square Collaborative Educational Mode of Higher Vocational ’Great Ideological and Political Classes‘ in Guangdong, Hong Kong and Macao under the Modernisation of Regional Vocational Education’ (23JDSZK085); Guangdong Provincial Vocational Schools Teaching Steering Committee for the Teaching of Commerce and Trade Specialities, Research on the Reform of Education and Teaching and Practices of the 2025 Project Research on the Construction of Civic and Political Demonstration Team for Innovative and Entrepreneurial Courses in Vocational Schools of Commerce and Trade, Project SM2025025.

References

[1] Guangdong Provincial Political Consultative Conference. Proposal on Promoting the High-Quality Development of Red Tourism in Our Province. 2025. Retrieved from [URL] [in Chinese].

[2] Wilson, M. Six views of embodied cognition. Psychonomic Bulletin & Review, 2002; 9(4), 625—636.

[3] Alyahya, M. & McLean, G. Examining Tourism Consumers’ Attitudes and the Role of Sensory Information in Virtual Reality Experiences of a Tourist Destination. Journal of Travel Research, 2022; 61(7), 1666–1681.

[4] Zhou X, Wong JWC and Wang SA. Memorable Tourism Experiences in Red Tourism: The Case of Jiangxi, China. Front. Psychol. 2022; 13:899144. DOI: 10.3389/fpsyg.2022.899144

[5] Xu T, Chen L, Tian M Z. A case study on Arishaty Resort in Hexigten World Geopark: Tourist’s Participation Motivation and benefit modes in geoparks. Resources & Industries, 2011,13(02):127—132. DOI: 10.13776/j.cnki.resourcesindustries. 2011.02.028.

[6] Zhao W H, Wang D. A Study on Tourist Participation Motivation in Virtual Brand Community: A Chinese Tourist Perspective. Enterprise Economy, 2011, 30(07):58—62. DOI: 10.13529/j.cnki.enterprise.economy.2011.07.050.
[7] Li Ling, Sui Haixia, Huang Ping. Research on the impact of tourism digital experience value on tourist citizenship behavior: An empirical analysis based on value co—creation theory. Resource Development & Market,2023,39(09):1239—1248.
[8] Li Ruyou, Shi Zhangyu. The influence mechanism of tourist value co—creation behavior on tourism satisfaction: A moderated mediation model. Arid Land Resources and Environment, 2022, 36(08):201—208. DOI: 10.13448/j.cnki.jalre.2022.218.
[9] G, S., Polisetty, A., Jha, R., & Keswani, S. Exploring innovation resistance in tourism: barriers to metaverse adoption among tourists. Cogent Business & Management, 2024;11(1).

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261082

Digital Management Innovation for Back-End Publishing Operations in Engineering Construction: A Lightweight Transformation Framework Based on Company A

Jing SONG[1]
Beijing Wuzi University, China
ORCID ID: Jing SONG https://orcid.org/0009-0004-8381-3740

Abstract. In the context of the digital economy and high-quality development of the engineering construction industry, digital transformation has entered a new stage of full-chain collaborative development. While digital practices in front-end design and construction have become mature, back-end support businesses represented by engineering publishing and document management still suffer from information asymmetry, process fragmentation, and data silos, which restrict the construction of a complete digital closed-loop system, Based on Business Process Reengineering (BPR), data element empowerment, and Task-Technology Fit (TTF) theory, this study takes Company A as an empirical case to construct a three-module lightweight digital management framework including information synchronization, visualized governance, and automatic data acquisition. This framework addresses operational inefficiencies in resource-constrained, non-standard scenarios without requiring large-scale investment. Empirical results show that the framework reduces the publishing cycle by 25%–48%, lowers the average consumable waste rate by 6.5 percentage points, improves data processing efficiency by more than 90%, and achieves 100% data accuracy. This study reveals that back-end digitalization is not simply system migration, but process reengineering, data empowerment, and lean operation optimization. For managers in resource-constrained back-end departments, this study provides a replicable, low-cost digital transformation path that avoids the implementation traps of large-scale ERP systems. The lightweight paradigm offers important theoretical and practical implications for digital management upgrading and high-quality development in the engineering construction industry.

Keywords. Engineering construction industry; back-end support services; digital management; lightweight transformation; business process reengineering; task-technology fit; case study

1. Introduction

1.1 Research Background and Problem Formulation

Driven by the 14th Five-Year Plan for Digital Economy Development, digital transformation has become a core strategy for quality improvement, efficiency

[1] Corresponding Author, Jing Song, Email: sj232232023@163.com

enhancement, and high-quality development in the engineering construction industry. The industry's digital construction has gradually shifted from single-point technology application to an in-depth reform of full-chain, systematic, and collaborative development [1]. At present, the engineering construction industry has formed mature technical systems and large-scale applications in front-end core business areas such as BIM technology, smart construction sites, and digital progress management [2]. However, as key links for engineering data implementation, achievement delivery, and compliant archiving, back-end support businesses represented by engineering publishing, document management, and printing services have long lagged in digital construction. The traditional operation mode relying on manual work and offline circulation is seriously inconsistent with the industry trend of in-depth digitalization and data-driven management [3,4].

As a typical back-end support business, engineering publishing is characterized by data intensity, high collaboration demands, strict timeliness and accuracy requirements, and non-standard production scenarios. Traditional operations lead to high transaction costs, information asymmetry, low resource allocation efficiency, and frequent quality defects, which directly restrict project delivery schedules and cost control [5]. In industrial practice, most enterprises hold a cognitive bias of "valuing front-end businesses while neglecting back-end businesses". Digital resources are concentrated on front-end businesses, leaving back-end business transformation lacking adaptive solutions and resource support.

Against this backdrop, post-2023 global research has formed two lightweight digital paradigms for budget-constrained SMEs, neither applicable to engineering printing scenarios.

The first paradigm is plug-and-play IIoT equipment digitalization proposed by McFarlane et al. [6], which adopts phased deployment without full hardware replacement. Yet it targets standardized machine lines and cannot adapt to construction's non-standard drawing orders and compliance rules. Anbalagan [7]'s lightweight digital twin research further proves mainstream industrial protocols are incompatible with multi-brand printers, breaking end-to-end linkage between workshop orders and equipment data.

The second paradigm is low-code/no-code (LCNC) office workflow platforms [8]. Multimodal foundation models reduce manual parsing of unstructured drawings [9], yet these tools merely solve cross-department information sharing and lack underlying equipment data docking capability.

Scholars diverge sharply on lightweight transformation priorities: some prioritize streamlined business processes with minimal hardware upgrades, while others regard equipment data collection as mandatory for closed-loop operation. The two paradigms each cover only one dimension, creating room for this study to integrate process collaboration and equipment perception.

Domestic lightweight drawing archive research lags behind international progress. Li [10] proposed a 2024 low-code archiving platform, but it only optimizes offline filing without real-time printer-order interconnection, failing to form a full-production closed loop.

Recent 2023–2024 studies prove digital success relies on matched technology and organizational culture rather than heavy investment; funding shortages and staff digital resistance remain top SME transformation barriers [11]. LCNC and foundation models lower technical thresholds and outperform traditional RPA [8,9], yet low digital maturity keeps paper-dependent workflows prevalent [12].

Collectively, existing lightweight frameworks suffer two shared flaws: manufacturing schemes only collect isolated equipment data and ignore cross-department coordination; office LCNC tools enable information synchronization but lack automatic equipment data capture. Though scholars uniformly advocate low-cost lightweight paths, no existing design addresses engineering publishing's triple constraints: heterogeneous multi-brand printers, strict drawing compliance, and variable batch project orders.

Beyond SME lightweight theories, traditional document transformation depends on full ERP deployment [5,13]. These heavyweight systems fit standardized mass production and require costly hardware renovation and dedicated IT teams. Additionally, mainstream ERP platforms lack native interfaces for multi-brand printing hardware, forcing expensive secondary customization to collect on-site production data.

Three distinct research strands leave an obvious gap: existing literature focuses solely on front-end construction digitalization, general SME lightweight schemes, or costly ERP document management. No study develops a targeted closed-loop lightweight framework that simultaneously satisfies engineering publishing's limited budgets, heterogeneous printing hardware, and non-standard order requirements.

In this study, "lightweight transformation" is defined from three management dimensions:

1) Technical lightness: no large-scale hardware replacement or complex system development;
2) Organizational lightness: no major organizational restructuring or staffing adjustment;
3) Cost lightness: low initial investment and short payback period.

In response to this identified research gap, this paper develops a management-oriented, low-cost and highly adaptable lightweight digital framework to improve operational efficiency and reduce internal management consumption.

1.2 Requirement Analysis and Research Ideas

This research responds to three-tier demands for construction back-office digitalization. Industrially, standardized, replicable back-end tools are required to close front-back digital divides and complete full-chain data circulation [3]. For enterprises, resource-limited publishing departments demand low-cost lightweight alternatives to expensive traditional digital systems [13]. Operationally, integrated information, visualized production and automatic data collection are essential to mitigate information friction and lower transaction costs [5].

Selecting Company A's publishing division as the case sample, this paper combines case analysis and quantitative measurement. It designs a lightweight framework matching customized, budget-restricted back-office scenarios, defines its implementation roadmap, and empirically validates its efficiency, cost and quality benefits to form a replicable back-office digital transformation scheme for construction enterprises.

2. Case Study Design

2.1 Case Selection Criteria

This paper adopts the case study method and selects Company A, a large representative enterprise in the engineering construction industry, as the research object. The case selection strictly follows the three principles of typicality, controllability and generalizability [14]. As a leading enterprise in the industry, the publishing service department of Company A is responsible for the printing and delivery of engineering drawings and technical documents of the whole company. Its business process covers the entire life cycle of engineering construction and production. The pain points existing before transformation are highly consistent with the common problems of large and medium-sized enterprises in the industry [4], which ensures the typicality of the case. Meanwhile, Company A did not replace production equipment, adjust staffing, or change the order structure and core business processes during the transformation. The core variables remained stable and the data were comparable, which can accurately verify the actual effect of the transformation scheme. The research conclusions can be applied to various back-end support businesses such as graphic publishing and document management [5], with extensive promotion value.

2.2 Data Sources and Variable Control

The research data are derived from first-hand operational data of Company A in Q1 2023 (pre-transformation) and Q1 2024 (post-transformation), with basically equal project quantity and printing workload. The data cover core operational indicators including the full publishing cycle, consumable usage, equipment data collection time, and task rework rate. The data are authentic and traceable, ensuring the objectivity and accuracy of empirical analysis [1]. To comply with academic norms and protect commercial secrets, enterprise information is anonymized in this study. To ensure research rigor, core variables such as equipment, personnel, and order scale are strictly controlled to eliminate external interference. The Hawthorne effect is excluded through long-term follow-up investigations (12 months of continuous observation) and cross-validation with departmental performance records, ensuring that the optimization of operational indicators results from the implementation of the lightweight digital transformation framework [1].

2.3 Evaluation Index System

Based on Hammer's business process reengineering theory [1], transaction cost theory, digital management theory [4], and Task-Technology Fit (TTF) theory [14], this study constructs a quantitative evaluation index system from three dimensions: efficiency, cost, and quality. In the efficiency dimension, three indicators (full publishing cycle, information synchronization time, and equipment data collection time) are selected to measure the improvement of operational efficiency. In the cost dimension, two indicators (consumable waste rate and labor cost saving rate) are used to evaluate the optimization of cost control. In the quality dimension, two indicators (data collection accuracy and task rework rate) are adopted to test the improvement of production quality. The index system balances scientificity and operability and can accurately reflect the practical value

of lightweight transformation [13]. Specific index definitions and calculation methods are consistent with the operational characteristics of engineering publishing business.

3. Case Enterprise Background and Core Pain Points

The publishing service department of Company A consists of a management layer and an execution layer. The traditional operation mode relies heavily on offline manual work, paper documents, oral communication, and handwritten records. Although the department has completed basic informatization transformation and accumulated continuous operational data, its digital foundation is relatively weak [4]. Its full business workflow includes five stages: demand reception, task scheduling, printing production, binding and finished-product delivery. Three critical operational bottlenecks exist:

First, cross-process information requires repeated manual retyping, causing data lag and frequent mistakes that hinder collaborative efficiency [5].

Second, there is no visual platform is available for production data. Real-time overall progress is inaccessible to supervisors, while frontline staff cannot obtain task updates promptly, raising communication and error-correction costs due to pervasive information asymmetry.

Third, all operational metrics of multi-brand printers are logged manually, creating heavy workloads and a 3%–5% recording error rate. These flawed records fail to support refined cost calculation and predictive equipment maintenance [5].

Meanwhile, the back-end business of Company A faces three major transformation constraints:

Production scenarios are highly non-standardized, and publishing requirements vary greatly across engineering projects, making traditional standardized digital systems unable to adapt to actual operational needs [13].

Inconsistent formats and transmission protocols between business and equipment data create high barriers to cross-source data integration [15];

The back-end department has a limited digital budget and insufficient technical capacity, making large-scale, high-investment digital transformation infeasible [14].

Against this background, a lightweight, low-threshold, high-benefit digital solution is urgently needed to break through the transformation dilemma [13]. It should also be noted that this single-case empirical setting inevitably restricts the generalizability of the proposed framework across organizations with disparate operational and organizational conditions.

4. Framework Design and Implementation of Digital Management

4.1 Overall Framework Design

Targeting the operational and digital transformation barriers in engineering drawing publishing, this study develops a lightweight three-module digital framework based on data element empowerment theory [15] and Task-Technology Fit theory [14]. Centered on information synchronization, visual monitoring, and automatic equipment data acquisition, the framework establishes a complete end-to-end digital workflow for task scheduling, on-site production and real-time supervision. Such task-technology matching avoids redundant, overdeveloped system functions and realizes the core

connotation of lightweight transformation. The full module interaction and data flow of this architecture are illustrated in Figure 1.

4.2 Core Module Design

The three modules address distinct operational bottlenecks and jointly enable closed-loop data operation. Their positioning and core functions are specified as follows.

4.2.1 Information Synchronization Module: Process Standardization and Integration

Drawing on BPR theory [1], the module unifies data standards [4] to support one-off data entry, cross-system automatic synchronization, real-time change alerts and full-process traceability. Archival modification records cut manual entry mistakes and information lag, reducing cross-department coordination costs [5]. Serving as the central data hub of the whole system, it auto-classifies uploaded publishing orders for full-process scheduling and synchronizes finished node status downstream.

4.2.2 Visual Management Module: Transparent Governance and Agile Decision-Making

Rooted in digital visualization theory [4], this module adopts a two-level dashboard as the unified display carrier. The global dashboard aggregates production progress, equipment utilization and logistics metrics, supporting multi-dimensional filtering and comparative analysis for managerial oversight. Frontline-specific dashboards deliver personalized task information, while workshop wall screens display real-time order, equipment and binding data. Visualized information reduces inter-post communication barriers and streamlines offline workflows [5].

4.2.3 Automatic Data Acquisition Module: Data Empowerment and Precise Management

Guided by data element value theory [15], customized interfaces across multi-brand printing equipment [2]. Serving as the bottom equipment perception layer, it automatically captures printer output, consumable usage and running status. After standardized data cleaning, equipment indicators are bound to production work orders to replace error-prone manual recording, supplying real-time credible data for precise cost calculation and predictive equipment maintenance [4,5].

4.2.4 Module Interaction Mechanism

Figure 1 marks nine key links (① – ⑨) to demonstrate the closed-loop data flow covering data collection, synchronization, visualization and on-site feedback. The hierarchical interaction logic of the three modules is sorted as follows:

1) Information Synchronization Module (Central Data Hub, Links ① – ③ & ⑤): External publishing requests are imported for demand classification and task scheduling (①). Triggered workflows dispatch tasks to on-site printing workshops (②), and real-time progress tracking records finished production nodes (③). Once binding and delivery are completed, the module automatically writes back full-process task records (⑤) and synchronizes all updated order data to the visual management module.

2) Automatic Data Acquisition Module (Equipment Perception Layer, Links ④ & ⑥): Continuous operational data are captured from various printing devices (④). Processed standardized data are uploaded to refresh real-time equipment displays on the visualization module (⑥).
3) Visual Management Module (Unified Presentation Layer, Links ⑦ – ⑨): Aggregate order progress and equipment data from the other two modules. It feeds updated task status back to the central hub to keep consistent information across all links (⑦); push integrated order, equipment and post-processing data to on-site workshop screens for frontline workers (⑧). Operators rely on visualized real-time data to arrange daily production and check work orders as required (⑨).

The above bidirectional data transmission builds an integrated closed-loop system. Each module undertakes independent layered functions while achieving full data connectivity and on-site production guidance.

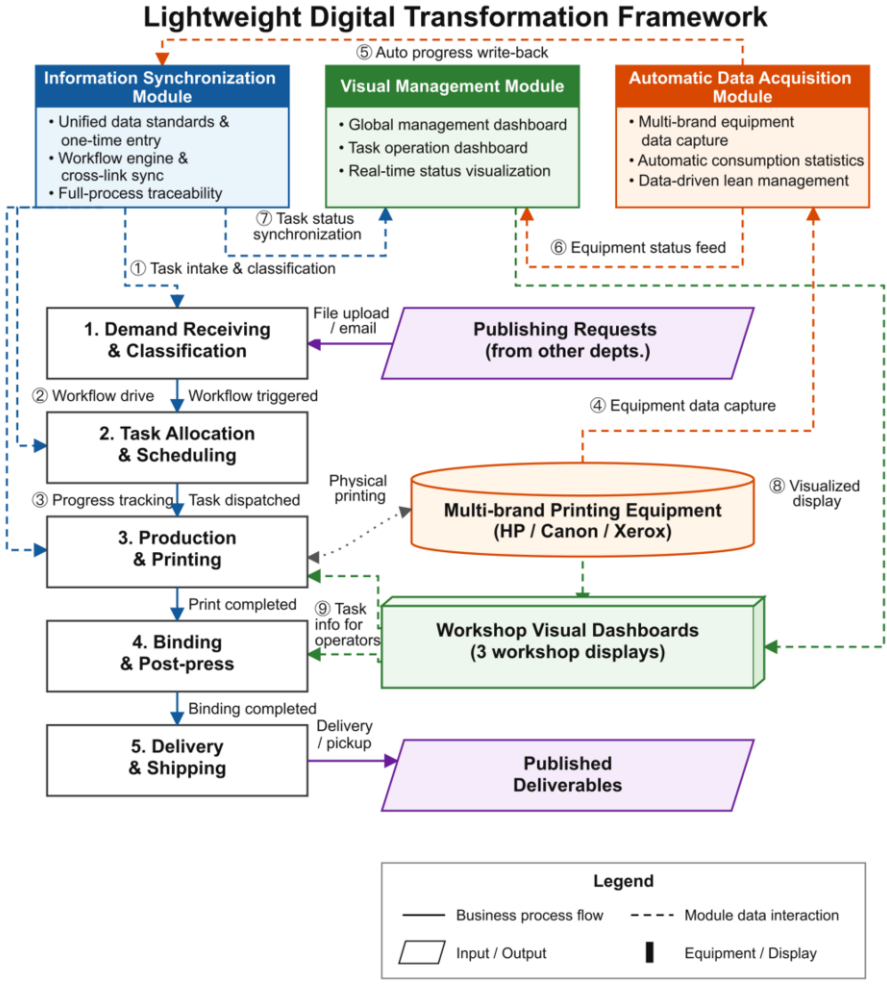


Figure 1. Data flow and module interaction diagram of the lightweight digital management framework

4.3 Implementation Path

This lightweight phased deployment scheme minimizes capital investment and transformation risks through four progressive stages, with clear rules for module launch sequence and cross-module interface debugging:

Stage 1: Full-process demand research and pain point sorting, defining standardized data indicators and functional requirements for each position [4].

Stage 2: Completing module development in the sequence of information synchronization, automatic data acquisition, and visual management, while debugging cross-module interfaces to ensure stable data transmission across system layers [2];

Stage 3: System trial operation, staff training, iterative optimization based on frontline feedback;

Stage 4: Official launch and normalized operation: monthly data review and quarterly function iteration to match evolving business needs [14].

This phased scheme fits back-end department operations with strong practicality [13].

5. Empirical Effect Analysis

Core variables are fully controlled with 12-month tracking to eliminate Hawthorne bias. Year-on-year Q1 datasets from 2023 and 2024 feature equivalent project quantities and printing loads, removing workload fluctuation interference. Table 1 quantifies multi-dimensional performance improvements brought by the lightweight framework.

Table 1. Comparison of core operational indicators before and after digital management transformation

Evaluation Dimension	Indicator Name	Pre-Transformation	Post-Transformation	Optimization Effect
Efficiency Dimension	Overall Publishing Cycle	46-96 h	24-72 h	Shortened by 25%-48%
	Equipment Data Collection Time	1-3 min/time	5-8 s / time	Improved by 86.7%-97.2%
	Information Synchronization Duration	2 min-2 h	Real-time response	100% synchronization efficiency improvement

Evaluation Dimension	Indicator Name	Pre-Transformation	Post-Transformation	Optimization Effect
Cost Dimension	Average Consumable Waste Rate	5%-20%	2%-10%	Decreased by an average of 6.5 percentage points
	Labor Communication Cost	20%-50% of working hours	Reduced to within 5%	An average saving of 83.3%
Quality Dimension	Data Collection Accuracy	95%-99%	100%	Error eliminated
	Task Rework Rate	8%-12%	2%-6%	Decreased by an average of 60%

5.1 Robustness Check and Indicator Interpretation

This paper eliminates three potential confounding factors to ensure the robustness of causal inference. First, no large-scale staff training or personnel turnover occurred during the observation period, avoiding interference from human resource changes. Second, the printing output of the two research quarters has a fluctuation of less than 2%, which excludes the impact of workload variation. Third, the same-quarter comparison effectively eliminates seasonal deviation.

The core operational improvements of the framework are summarized as follows. Automated data entry and cross-team information synchronization release human resources from repetitive work to high-value links including quality control and customer service, without downsizing. In addition, real-time equipment monitoring and precise material allocation reduce subjective overstock and consumable waste. Overall, the proposed framework achieves coordinated optimization of collaboration efficiency, operational cost, and production quality.

6. Conclusions and Future Prospects

6.1 Research Conclusions and Managerial Implications

Targeting prominent operational bottlenecks within engineering drawing publishing, this paper develops a lightweight three-module digital framework. Empirical verification at Company A verifies its comprehensive gains in operational efficiency, cost management and production quality for cost-limited internal administrative departments. The

framework fills existing gaps in the digital transformation research of construction drawing management and delivers a low-cost, replicable implementation solution.

First, successful back-office digital transformation hinges on business process reengineering and data integration instead of complex customized development; technical solutions need to align with actual workflow logic to unlock digital value[15]. Second, incremental lightweight transformation better accommodates limited administrative budgets, delivering tangible operational outcomes with modest capital and labor input. Third, enterprises ought to avoid overemphasis on front-end digital construction alone. Balanced digital investment across front and back offices is essential to achieve end-to-end data connectivity and cross-department collaboration [3].

6.2 Applicability Boundaries and Research Limitations

6.2.1 Applicable Scenarios

This study relies on single-case empirical evidence. Accordingly, we analyze how the lightweight framework fits varied organizational environments, and put forward differentiated expansion and transformation strategies for different construction organizations:

For highly applicable scenarios, the framework achieves optimal efficiency for medium and large construction enterprises that possess basic digital infrastructure, suffer cross-departmental data silos, deploy multi-brand printing hardware, and undertake stable batch drawing printing tasks. No additional customized development is required; the native three-module architecture can be deployed directly to realize full-process closed-loop data operation.

In case of conditionally applicable scenarios, two types of organizations fall into this category, each with clear targeted expansion modifications:

- Small-scale construction enterprises: Complete unification of enterprise data standards and full networking of printing equipment is a mandatory precondition before deployment. Adding lightweight standardized data templates to reduce data sorting costs without large-scale reconstruction.
- Multi-branch enterprise groups: The native single-workshop data aggregation logic cannot support cross-site unified supervision. Supplement a multi-site data aggregation sub-module attached to the visualization module, to synchronize workshop production data onto a unified top management dashboard.
- Engineering subcontractors with scattered small-batch orders: Special printing devices for customized drawings lack built-in adaptive interfaces. Secondary development of compatible lightweight interface plugins can be carried out to realize automatic data collection for non-mainstream printing equipment.
- State-owned construction enterprises with strict offline compliance approval procedures: Connect independent middleware between the data synchronization module and corporate OA systems to link digital production and offline approval workflows.

When it comes to low-adaptability scenarios, fully offline micro drawing workshops without networked printing devices and standardized order management cannot deploy the framework directly. Complete digital reconstruction of basic data and equipment

networking is required in advance, which raises upfront transformation costs and erodes the lightweight cost advantage of the framework.

6.2.2 *Multi-Dimensional Research Limitations and Targeted Mitigation*

1) Single-case restriction: All empirical data originate solely from Company A's publishing division. Differences in enterprise scale, culture and internal management mean the measured efficiency, cost and quality improvements cannot be universally replicated across construction back-office departments. To enhance generalizability, future work will carry out multi-case comparisons covering enterprises of diverse sizes and branches, with customized framework versions developed for differentiated operational scenarios
2) Organizational digital resistance: The model assumes full employee acceptance of digital workflows, ignoring two core barriers: veteran frontline staff's reliance on paper operations, and middle managers' resistance to transparent offline information control. Such resistance reduces system utilization, delays real-time data synchronization and impairs long-term framework value. This study mitigates conflicts via human-centered design: one-click data entry reduces manual labor; hierarchical dashboards distribute access rights rationally; a phased online-offline hybrid mode lowers adaptation pressure for digitally inexperienced staff.
3) Limited technical scalability: The architecture is customized for engineering drawing printing only, lacking native modules for procurement, document archiving and other auxiliary businesses. Its built-in interfaces merely support mainstream printers, requiring costly secondary development for extended business or non-standard printing hardware. All core modules reserve standard open interfaces to address this flaw; follow-up iterations will integrate a universal interface template library to shorten custom development cycles and cut extra R&D expenditure.
4) Short-term observation window: Performance assessment relies on short-term tracking data and fails to capture long-term hidden operation costs, including routine upgrade labor, storage overhead from accumulated logs and repeated manual logic debugging amid changing project requirements. These burdens may erode the lightweight scheme's cost advantages over time. Subsequent updates will embed automatic data archiving and compression tools, plus configurable adaptive workflow templates to reduce storage pressure and regular manual maintenance work.

6.3 Future Prospects

Relying on the operational dataset accumulated by this lightweight framework [13], follow-up work can build AI modules for demand forecasting, equipment fault warning and intelligent scheduling to realize predictive active operation [2,16]. Multimodal foundation models and big data analytics will further advance semi-unmanned back-office production and release data element value [15].

In addition, future research can conduct multi-case comparative studies covering construction firms of varied scales and digital maturity to revise the framework and improve its universality [14]. Meanwhile, unified industrial digital standards for engineering publishing will also be formulated [4] to popularize lightweight

transformation schemes and narrow the front-back digital divide for full-chain industrial upgrading [3,13].

References

[1] Y. Chen, Z. Liu, and M. Wang, "Digital transformation and operational efficiency improvement of engineering publishing business," in 2021 Digital Management Innovation Conference. IOS Press, 2021, pp. 78–85.

[2] M. Bilal, T. Zayed, and M. Alshawi, "Digital transformation in construction: A systematic review," Automation in Construction, 2019 Sep.; vol. 106, p. 102825 , doi:10.1016/j.autcon.2019.102825.

[3] G. Q. Wang, "Research on lightweight digital collaborative management platform for metro engineering project design," Construction Science and Technology, 2024 Sep.; vol. 24, no. 18, pp. 101–103.

[4] L. L. Ren, Q. Zhang, and J. M. Guo, "Research on digital transformation mechanism of incumbent enterprises based on dynamic capability theory," Economic Forum, 2025 Sep.; vol. 48, no. 9, pp. 72–80.

[5] Y. Zhou, J. Lyu, and L. Li, "Can digital transformation enhance labor productivity in enterprises: An analysis from the perspective of business process reengineering," PLOS ONE, 2025 Jun.; vol. 20, no. 6, pp. 1–21, doi:10.1371/journal.pone.0325484.

[6] D. McFarlane, S. D. L. Silva, and R. Batista, "Progress Towards Low-Cost Industrial Digitalisation for SMEs," IFAC-PapersOnLine, 2024; vol. 57, no. 11, pp. 896–901, doi:10.1016/j.ifacol.2024.09.149.

[7] A. Anbalagan, "Lightweight CNC digital process twin framework: IIoT integration with open62541 OPC UA protocol," Production & Manufacturing Research, 2025; vol. 13, no. 1, pp. 1–22, doi:10.1080/21693277.2025.2544981.

[8] M. O. Ajimati, N. Carroll, and M. Maher, "Adoption of low-code and no-code development: A systematic literature review and future research agenda," Journal of Systems and Software, 2024 Apr.; vol. 210, p. 112873, doi:10.1016/j.jss.2024.112873.

[9] M. Wornow, A. Narayan, K. Opsahl-Ong, Q. McIntyre, N. H. Shah, and C. Ré, "Automating the enterprise with foundation models," Proceedings of the VLDB Endowment, 2024 May; vol. 17, no. 9, pp. 2341–2354, doi:10.14778/3681954.3681964.

[10] X. Li, "Lightweight low-code management platform for construction drawing archives under limited IT budgets," Journal of Construction Management, 2024 Apr.; vol. 26, no. 4, pp. 56–63, doi:10.1234/jcm.2024.26.4.56.

[11] A. Kargas, E. Gialeris, F. Komisopoulos, A. Lymperopoulos, and I. Salmon, "Digital maturity and digital transformation strategy among Greek small and medium enterprises," Administrative Sciences, 2023 Nov.; vol. 13, no. 11, p. 236, doi:10.3390/admsci13110236.

[12] S. Sternad Zabukovsek, S. Jordan, and S. Bobek, "Managing document management systems' life cycle in relation to an organization's maturity for digital transformation," Sustainability, 2023 Nov.; vol. 15, no. 21, p. 15212, doi:10.3390/su152115212.

[13] Y. Ma, "Exploration and application of digitalization in engineering document management," Technology and Information, 2023 Mar.; vol. 42, no. 3, pp. 63–68.

[14] D. L. Goodhue, "Task-technology fit and individual performance," MIS Quarterly, 1995 Jun.; vol. 19, no. 2, pp. 173–192, doi:10.2307/249659.

[15] H. Tian and S. X. Wang, "Mechanism and path of value creation of prefabricated buildings based on data elements," Journal of Engineering Studies, 2023 Feb.; vol. 15, no. 1, pp. 62–69.

[16] IAARC, "A systematic data-driven framework for integrating AI and robotics in construction," in Proc. CSCE CRC 2025, Toronto, Canada, 2026 Feb.; pp. 1–10.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261083

Demand Forecasting of Logistics Management Talents in Guangxi under Digital-Intelligent Transformation

Yanni LIANG[1], Xiaoqin WANG, Shanshan ZHENG and Juan HUANG
Beibu Gulf University, Qinzhou, Guangxi 535011, China

Abstract. In the digital era, the logistics industry has a significant demand for digital-intelligent talents. Addressing the structural contradiction between supply and demand of logistics talents in Guangxi, this paper analyzes data from 2021 to 2025 across three dimensions—region, position, and education level. Correlation screening identifies the number of newly registered logistics enterprises in Guangxi as the indicator with the strongest observed association. Given the longer enterprise-registration series and limited paired outcome data, a two-stage forecasting model is constructed: ARIMA first forecasts newly registered logistics enterprises for 2026-2030, and linear regression then maps these values to graduate demand. The analysis results show: a high regional concentration in Nanning; distinct job differentiation, with average supply-demand ratios of 2.85, 3.9 and 3.89 for logistics/warehousing, transportation and procurement categories respectively, all indicating oversupply, while the production and manufacturing category has a ratio of only 0.37, indicating undersupply; and prominent educational mismatch, with a supply-demand ratio of 10.93 for bachelor's degrees in 2025, compared to only 0.9 for associate degrees. According to forecasts, demand will grow steadily over the next five years, with an average annual increase of 10.07%, mainly at the associate degree level. An additional 1,000 newly registered logistics enterprises are associated with an estimated 484 more graduates demanded. Based on this, four countermeasures are proposed: training system, integration of industry and education, regional balance, and out-of-region employment, providing a reference for the reform of logistics talent training in Guangxi. Given the five-year paired outcome series, the forecasts are indicative planning scenarios.

Keywords. Logistics management, digital-intelligent transformation, talent demand forecasting, Guangxi

1. Introduction

Talent aligned with regional industry is central to logistics development [1]. Chinese studies forecast provincial logistics talent demand using regression [2,3], fuzzy combination models [4], and GM (1,1) for small samples [5]. Human-capital theory views education and training as productivity-enhancing investments [6], while labor-market matching stresses alignment between supplied skills and job requirements. International research identifies digital-skill shortages as an Industry 4.0 workforce challenge [7] and supports experiential Logistics 4.0 training [8]. Recent logistics studies further show that digitization readiness depends on functional training and organizational

[1] Corresponding Author: Yanni Liang, yannibeijing@163.com.

support [9], that firms in emerging economies face gaps between current and required qualifications [10], and that Logistics 4.0 technologies reshape operator skills and training needs [11]. However, these technological changes have not been adequately incorporated into higher education, leading to a growing disconnect between academic programs and corporate practice. Given the accelerating digital and intelligent transformation of the logistics industry and the widespread adoption of technologies such as automated sorting and drone delivery, this disconnect has become more pronounced. This has given rise to a structural contradiction characterized by a shortage of technical talents in fields such as big data and AI, as well as high turnover rates among frontline operational staff. The Action Plan for Accelerating the Cultivation of Digital Talents to Support the Development of the Digital Economy (2024–2026) explicitly calls for increasing the supply of digital talents. Taking enterprises in Guangxi as an example, this paper forecasts the demand for logistics management professionals, aiming to provide a reference for the reform of logistics management talent cultivation in Guangxi's secondary and tertiary institutions.

2. Supply and Demand Analysis of Graduates in Guangxi's Logistics Talent Market

The logistics industry serves as an important channel for employment. With the deepening of RCEP and the improvement of Guangxi's transportation network, the demand for relevant talents continues to grow. According to statistics from the Guangxi Talent Network, the number of logistics/warehousing talent demanded in Guangxi in 2025 reached 11,140 (of which graduates from secondary and tertiary institutions accounted for 6.7%), representing a year-on-year increase of 25.62%. A comparative analysis of the supply and demand data for graduates from secondary and tertiary institutions in Guangxi from 2021 to 2025 reveals the following: (1) Significant regional imbalance: Educational resources and leading enterprises are highly concentrated in Nanning, where the proportions of both job-seeking graduates and talent demand exceed those of other cities by more than tenfold. (2) Severe job structure differentiation: The annual supply-demand ratios for the mainstream positions in logistics management—logistics/warehousing and transportation—are 2.85 and 3.9. Among strongly related positions, the manufacturing/production management category has a ratio of only 0.37, indicating a shortage of supply, while the procurement category faces intense competition with a ratio of 3.89. (3) Pronounced educational mismatch: In 2025, the supply-demand ratios for bachelor's degree holders and postgraduate degree holders (master's and above) stood at 10.93 and 18.16, whereas the ratios for associate degree holders and those with secondary technical education or below were 0.9 and 0.37. This indicates that the market urgently needs applied skilled talents rather than merely those with higher academic degrees. Therefore, logistics management talent cultivation in Guangxi's secondary and tertiary institutions must promptly align with market demand, optimize the professional structure, and strengthen the supply of practical skills.

3. Forecast of Logistics Talent Demand in Guangxi

3.1. Basic Data Analysis

Taking the number of graduates from secondary and tertiary institutions demanded by Guangxi's transportation, storage, and postal industry from 2021 to 2025 as the dependent variable (*Y*), and drawing on relevant studies on the influencing factors of talent demand, we collected data on various indicators for the same period to construct the foundational dataset for the logistics talent demand forecasting model (Table 1). Graduate-demand data are publicly available only for 2021-2025; therefore, correlation and Ordinary Least Squares (OLS) use five paired annual *X*-*Y* observations, while the 2006-2025 *X* series (20 observations) is used only for ARIMA forecasting. Missing 2025 e-commerce sales and 2024-2025 smart-logistics market values were estimated by compound-growth extrapolation and used only for preliminary indicator screening.

(1) Compound growth rate extrapolation method

$$c = \left(V_{t+m} / V_t\right)^{1/m} - 1 \tag{1}$$

$$V_{t+j} = V_t \left(1 + c\right)^j \tag{2}$$

Where: V_t is the observed value in year t; m is the number of intervals; c is the compound growth rate; j is the number of extrapolation years.

(2) Correlation analysis based on Pearson correlation coefficient

Pearson's coefficient screened linear associations with graduate demand. Absolute coefficients of 0.8-1.0 were classified as extremely strong, 0.6-0.8 as strong, 0.4-0.6 as moderate, 0.2-0.4 as weak, and below 0.2 as negligible.

$$r = \sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right) \Bigg/ \left(\sqrt{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}\sqrt{\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2}\right), \quad -1 \le r \le 1 \tag{3}$$

Where: r is the correlation coefficient; X_i is the independent variable; Y_i is the dependent variable; $\bar{X}$ is the mean of the independent variable; $\bar{Y}$ is the mean of the dependent variable.

Only five paired *X*-*Y* observations are available; multiple regression would exhaust degrees of freedom, and additional unpaired X years cannot be used in OLS. The number of newly registered logistics enterprises, which has the strongest screening correlation with demand (r=1.00, though this value is sensitive to the small sample), is therefore used in a parsimonious simple regression. Variable screening and significance tests are exploratory and do not establish causality.

Table 1. Basic data

Indicator \ Year	2021	2022	2023	2024	2025	Correlation coefficient *r*	Correlation judgment
Number of graduates from secondary and tertiary institutions demanded by	1186	849	721	1964	1094	/	/

Indicator \ Year	2021	2022	2023	2024	2025	Correlation coefficient *r*	Correlation judgment
Guangxi's transportation, storage and postal industry (persons)							
Number of newly registered logistics enterprises in Guangxi (enterprises)	7917	7287	7047	9577	7617	1.00	Extremely strong
Total import and export volume of Guangxi with ASEAN countries (100 million yuan)	2821.17	2811.13	3394.43	3978.19	4292.16	0.44	Moderate
Guangxi railway operating mileage at year end (km)	5216	5337	5672	5866	5947	0.37	
Guangxi annual total freight volume (100 million tons)	21.62	21.33	22.84	23.71	24.78	0.36	
Guangxi postal business volume (100 million yuan)	165.96	175.96	196.48	224.88	251.56	0.36	
China's smart logistics market size (100 million yuan)	6477	6995	7903	*8913.00*	*9906.64*	0.34	
Value added of Guangxi's transportation, storage and postal industry (100 million yuan)	998.5	1003.3	1075.3	1115.38	1165.22	0.33	Weak
Guangxi GDP (100 million yuan)	24740.86	26300.87	27202.39	28649.4	29727.45	0.32	
Guangxi port container throughput (10,000 TEUs)	719.82	827	957.35	1080.04	1197.18	0.32	
Guangxi total import and export volume of goods (100 million yuan)	5930.63	6603.53	6936.49	7563.89	8192.62	0.29	
Guangxi highway mileage at year end (km)	160600	172400	183600	187000	188500	0.23	
Guangxi e-commerce sales (100 million yuan)	2616.6	3703.8	4022.6	4789.4	*5874.17*	0.23	

Note: The data come from Guangxi Talent Network, Tianyancha, Guangxi Bureau of Statistics, Guangxi Postal Administration, Guanyan Report Network, and the National Bureau of Statistics; China's smart logistics market size in 2019 and 2023 was 488.5 billion yuan and 790.3 billion yuan, respectively; Guangxi's e-commerce sales in 2020 was 211.65 billion yuan; italicized values are extrapolated.

3.2. Simple Linear Regression Model

OLS was implemented in SPSSAU using the five calendar-year-matched observations of the number of graduates from secondary and tertiary institutions demanded by Guangxi's transportation, storage, and postal industry (Y) and the number of newly registered logistics enterprises in Guangxi (X). The procedure estimates the regression equation, R^2, and significance tests. $\hat{Y}$ denotes predicted demand, a is the intercept term, b is the slope (regression coefficient), and $\bar{X}$ and $\bar{Y}$ are the means of variables X and Y, respectively.

$$\hat{Y} = a + bX \tag{4}$$

$$b = \sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right) \Big/ \sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 \tag{5}$$

$$a = \bar{Y} - b\bar{X} \tag{6}$$

3.3. Results Analysis

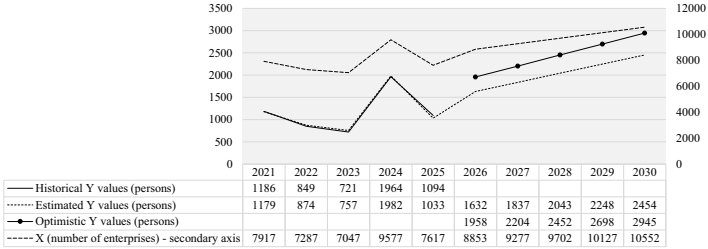
Table 2 reports the in-sample results. R^2=0.994 and the F-test is significant (p<0.001), yielding Y=−2653.264+0.484X. With one predictor, VIF=1.000 by construction. The Durbin-Watson statistic is 1.281, but n=5 gives little diagnostic power. The slope (p<0.001) implies that 1,000 additional newly registered logistics enterprises correspond to an estimated 484 more graduates demanded. Because r=1.00 and R^2 are sensitive to individual observations in such a small sample, they indicate association rather than causality or assured out-of-sample accuracy.

Table 2. Linear regression analysis results (n=5)

	Unstandardized coefficients		Standardized coefficients	t	p	Collinearity diagnostics	
	B	**Std. Error**	***Beta***			**VIF**	**Tolerance**
Constant	-2653.264	177.185	-	-14.975	0.001**	-	-
X	0.484	0.022	0.997	21.675	0.000**	1.000	1.000
R^2	0.994						
Adjusted R^2	0.992						
F	F(1,3)=469.806, p=0.000						
D-W value	1.281						

Note: ** indicates p<0.01.

An ARIMA(0,1,1) model uses 20 annual observations of newly registered logistics enterprises (X, 2006-2025) to forecast X for 2026-2030; the five-point Y series is too short for direct ARIMA estimation. Forecast X values are substituted into the existing regression to obtain Y. Figure 1 shows that the historical Y values from 2021 to 2025 fluctuated significantly, with a sharp increase in 2024 and a decline in 2025, reflecting instability in industry demand, yet the overall trend remained consistent with X. The predicted values indicate that after a short-term drop in 2025, the demand will enter a stable growth path. The forecast demand grows by about 10.07% annually during 2026-2030. The 20% upward case is an illustrative sensitivity scenario reflecting favorable Pinglu Canal and ASEAN conditions, not a statistically estimated forecast. The two-stage design limits parameters but does not remove small-sample uncertainty.



	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030
Historical Y values (persons)	1186	849	721	1964	1094					
Estimated Y values (persons)	1179	874	757	1982	1033	1632	1837	2043	2248	2454
Optimistic Y values (persons)						1958	2204	2452	2698	2945
X (number of enterprises) - secondary axis	7917	7287	7047	9577	7617	8853	9277	9702	10127	10552

Figure 1. Demand for graduates from Guangxi's transportation, storage and postal industry and number of newly registered logistics enterprises, 2021–2030.

Note: Values of X for 2021–2025 are historical data; values for 2026–2030 are forecast data.

For simplicity of calculation, based on the forecasted regional graduate demand for 2027 (1,837 persons), the demand for graduates at different educational levels in Figure 2 for 2027 was derived using the average proportions of different educational levels from 2021–2025 as recorded by the Guangxi Talent Network. The demand for graduates in each city was derived using the average proportion of regional demand to total regional demand from 2021–2025. Figure 2(a) indicates that logistics talent cultivation should emphasize application and practice, particularly the development of digital and intelligent technical skills, because industry demand is highly concentrated at the associate degree level (accounting for 74.52%), and positions with no educational requirement account for 12.47%, highlighting the skill-oriented characteristics of employment. Figure 2(b) shows that regional demand is highly concentrated in Nanning (accounting for 78.55%). Nanning's hub role should be leveraged to facilitate regional coordination and talent mobility through industrial chain collaboration and gradient transfer. Meanwhile, training resources should be proactively directed to Qinzhou, Hengzhou, and Lingshan to meet the growing demand for port-logistics and multimodal transport professionals driven by the Pinglu Canal project.

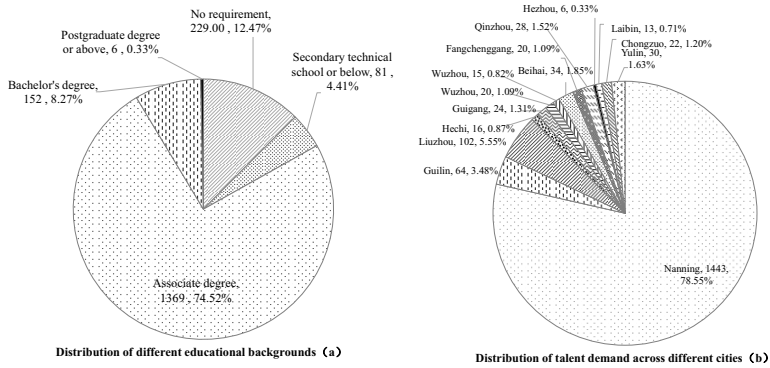


Figure 2. Demand for graduates from Guangxi's transportation, storage and postal industry by educational level and by city, 2027.

4. Strategies for Cultivating Digital-Intelligent Logistics Management Talents in Guangxi

First, optimize training: add modules/practice for production shortages, refine procurement/competitive courses, upgrade mainstream; enhance associate-level applied training with digital integration and more practice, shifting from degree to skills.

Second, deepen integration: co-build internship bases/industrial colleges aligned with job standards; leverage Nanning to build a community with early-warning mechanism to adjust programs; collaborate with manufacturers to train production logistics talent.

Third, promote balance: leverage Nanning's hub to radiate resources; boost policy/infrastructure in Qinzhou, Beihai, Chongzuo to create jobs; provide targeted training for Pinglu Canal port/logistics; offer housing/career incentives to attract graduates to neighboring cities.

Fourth, expand out-of-region employment to ease oversupply: target Pearl River Delta, Yangtze Delta, Greater Bay Area for "tech+ops" talent; partner with top firms (SF,

JD, Cainiao, etc.) to place graduates in analytics, smart warehousing, cross-border ops, supply chain planning; use job fairs and alumni platforms to improve placement.

5. Discussion

Unlike earlier province-level studies focused on aggregate forecasts or economic indicators [2-5], this study disaggregates Guangxi's mismatch by region, occupation, and education. Its applied, associate-level demand aligns with international evidence on digital-skill change, practice-oriented Logistics 4.0 training, manpower readiness, and qualification gaps [7-11], while Nanning's concentration and production-logistics shortages reflect Guangxi's local industrial structure.

The forecasts are exploratory and should be treated as indicative rather than definitive. Five paired Y observations cannot support a meaningful holdout or rolling validation, and extending X alone does not increase the regression sample. High r and R^2 may reflect common movement, variable selection, or influential years. The regression provides no external validation or prediction interval, while compound-growth estimates and the illustrative 20% sensitivity case add uncertainty. Results should therefore guide planning rather than be interpreted as causal evidence or confirmed policy effects.

6. Conclusion

Against the backdrop of digital-intelligent transformation, Guangxi's logistics talent market faces prominent structural contradictions: regionally concentrated in Nanning; oversupply in logistics/warehousing, transportation, and procurement, while manufacturing/production faces a shortage; excess of bachelor's degree holders and a shortage of associate degree holders, with increasing demand for digital-intelligent skills. It is forecast that from 2026 to 2030, the demand for graduates will grow steadily at an average annual rate of 10.07%, rising from 1,632 to 2,454 persons, showing an accelerating upward trend. In 2027, demand will be highly concentrated at the associate degree level (74.52%) and in Nanning (78.55%). Accordingly, the findings suggest four planning options: optimize the training system by strengthening courses for shortage positions and enhancing digital skills; deepen industry-education integration by reinforcing university-enterprise collaboration and dynamic monitoring; guide regional balance by directing talent toward the Pinglu Canal area and other potential cities; expand out-of-region employment by sending talents to developed regions. These are planning options rather than confirmed policy effects and require further empirical evaluation. Future research should obtain a longer matched X-Y series, report prediction intervals and external or rolling validation, and compare small-sample alternatives such as grey or Bayesian forecasting.

7. Funding

This research was funded by Guangxi Higher Education Undergraduate Teaching Reform Projects: "'Digital Intelligence + Industry-Education Integration': Innovation and Reform of Cultivating Compound Logistics Management Talents in Local Universities

Based on the OBE Concept" (2024JGA261); "Research and Practice on the Cultivation Model of New Quality Productive Forces Talents in Logistics Management under the Context of the Pinglu Canal Construction" (2024JGA265); and "Digital Empowerment ·Virtual-Real Symbiosis: Innovation and Practice of the 'AI + Industry-Education Integration' Teaching Model for the Course 'Management' under the New Liberal Arts Context" (2025JGA291). It was also funded by the Key Topic of Guangxi Science and Technology Think Tank (Guike Association [2022] P-53). The authors also thank the editor and the anonymous reviewers for their valuable comments.

References

[1] Ning C. Research on the Demand for Logistics Talents in the Shenyang Economic Zone. Modern Education Management. 2011 Feb;(21):54-56, doi: 10.16697/j.cnki.xdjygl.2011.02.034

[2] Li XL, Yan XQ, Wu AP. Research on forecasting the demand for logistics talents in Guangdong Province. China Storage & Transport. 2016 Dec;(12):130-132, doi: 10.16301/j.cnki.cn12-1204/f.2016.12.051

[3] Tang YZ. Forecast and analysis of the demand for logistics skilled talents in the logistics industry in Zhejiang Province. China Market. 2018 Jun;(18):167-169, doi: 10.13939/j.cnki.zgsc.2018.18.167

[4] Guo ZX, Cheng T, Zhang YH. Forecasting the Demand for Logistics Talents in Hebei Province Based on Fuzzy Combination Forecasting. Journal of Hebei Agricultural University (Social Sciences). 2019 Jan;21(1):125-129, doi: 10.13320/j.cnki.jauhe.2019.0021

[5] Jiang N. Forecast of Rural Logistics Talent Demand in Nanling County, Wuhu City - Based on the Grey GM(1,1) Model. Modern Business. 2021 Feb;(4):61-63, doi: 10.14097/j.cnki.5392/2021.04.018

[6] Becker GS. Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education. 3rd ed. Chicago: University of Chicago Press; 1993.

[7] Ozkan-Ozen YD, Kazancoglu Y. Analysing workforce development challenges in the Industry 4.0. International Journal of Manpower. 2022 May;43(2):310-333, doi: 10.1108/IJM-03-2021-0167

[8] Belmonte LM, Segura E, Lopez de la Rosa F, Gomez-Sirvent JL, Fernandez-Caballero A, Morales R. Training industrial engineers in Logistics 4.0. Computers & Industrial Engineering. 2023 October;184:109550, doi: 10.1016/j.cie.2023.109550

[9] Gupta A, Singh RK, Gupta S. Developing human resource for the digitization of logistics operations: readiness index framework. International Journal of Manpower. 2022 May;43(2):355-379, doi: 10.1108/IJM-03-2021-0175

[10] Sezer IC, Sorkun MF. An investigation on logistics firms' human resources qualifications in transition to Industry 4.0: An insight from Türkiye. Procedia Computer Science. 2024 March;232:2580-2587, doi: 10.1016/j.procs.2024.02.076

[11] Lagorio A, Di Pasquale V, Cimini C, Miranda S, Pinto R. Augmented Reality in Logistics 4.0: implications for the human work. IFAC-PapersOnLine. 2022 October;55(10):329-334, doi: 10.1016/j.ifacol.2022.09.408

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261084

Research on the Application Practice Methods of Data Intellectual Property Protection

Qin LEI[a,1] and Yajuan TIAN[b]
[a]*Sichuan University Library, China*
[b] *National Science Library (Chengdu), Chinese Academy of Sciences, Chengdu, China*
ORCiD ID: Qin LEI https://orcid.org/0000-0002-5338-1949
Yajuan TIAN https://orcid.org/0000-0002-0467-2432

Abstract. Data is gradually becoming a crucial focus of high-quality economic development, making it imperative to reasonably define the legitimate rights and interests of all parties involved in data production, circulation, and utilization, and to explore the establishment of appropriate data protection rules. This study employs online research, literature review, and comparative analysis to examine the relevant laws and regulations on data empowerment issued by some countries and regions, focusing on these aspects: the applicability and limitations of laws related to data intellectual property (DIP) protection, and the current application practices and methods of DIP. The results provide a systematic reference for understanding the existing legal frameworks and practical approaches to DIP protection.

Keywords. Big data, Intellectual property, Data protection, Data property

1. Introduction

After agricultural and industrial economies, the digital economy has become a new economic form today. The commercial value of data is increasingly prominent. With the development of artificial intelligence (AI), 5G, and the Internet of Things (IoT), the role of data in technological and industrial transformation has gained growing significance [1]. Intellectual property (IP) is a crucial means for technological innovators to protect their legitimate rights and interests, and is the core foundation for maintaining competitive advantages [2]. IP system is in line with the data property system in terms of incentivizing innovation and improving efficiency.

Therefore, many countries are promoting the implementation of data intellectual property (DIP) protection projects, exploring legislative research on DIP, accelerating the construction of DIP protection rules, and better promoting the rational data flow, effective DIP protection, and full utilization of data factor on the basis of protecting personal privacy and national data security. The theoretical research and practical exploration of DIP are still in their early stages. At the statutory level, there is currently

[1] Corresponding Author, Qin Lei, Sichuan University Library, China; Email: leiqin@scu.edu.cn. The work was supported by the Research Projects of the Library Society of China "Research on the Patent Layout Information Service Model of University Libraries for Technological Innovation Subjects" (Grant No. 2026LSCFZZYB0089).

no formal definition of DIP. In this study, DIP refers to rights with intellectual property attributes arising from the creation, development, and utilization of data. The protection of DIP requires the joint participation of multiple fields and institutions to build a diversified protection system for different types of data. However, due to the complexity involved in data and the diverse nature of data protection, there is controversy over how to establish a comprehensive data property protection system [3].

Data protection also involves a balance between protecting the legitimate rights and interests of data owners and ensuring the rational flow of data. Although the current judicial practice applies the existing IP legal protection system to empower certain type of data, the limitations are also clearly evident.

2. Literature review

2.1. IP Protection Characteristics of Data

Based on the characteristics of data, some researchers have studied the rationality and feasibility of using existing IP laws to protect certain data. Feng pointed out that data has property attributes and should be protected by law [4]. Data that meets the protection conditions of IP laws can be included in the scope of IP protection. Other data can be included in the protection of the anti-unfair competition law when there are property interests or competitive interests that deserve legal protection.

Based on the relevant data law system in South Korea, Rae studied the theoretical background, relevant laws, data transactions, and data IP laws for reestablishing the concept of data ownership [5]. Kim proposed legislative measures to promote the balance between big data utilization and information protection based on the legal status of data as the object of rights and obligations and the legal relationship between data contract parties [6]. Suo and Pan believe that commercial data can use IP rights as a reference to legislate for data property rights and connect them with other rights [7]. Shan proposed that China should incorporate data into existing IP objects [8].

In addition, the relevant provisions in the WIPO Copyright Treaty (WCT) also state that intellectual creations composed of the selection or arrangement of compilations of data or other material should be protected (Article 5) [9]. The Agreement on Trade Related Aspects of Intellectual Property Rights (TRIPS) also mentions the protection of original data compilations (Article 10(2)) [10]. With the deepening development of large language models (LLMs), the unauthorized use of copyrighted data to train LLMs is also a very cutting-edge and complex legal and ethical issue. EU AI Act passed in 2024 indirectly involves the protection mode of training data [11], and the Interim Measures for the Administration of Generative Artificial Intelligence Services released in 2023 in China, stipulate that training data should meet the requirements of legality shall not infringe on intellectual property rights [12].

Furthermore, Tim Berners Lee's Open Data Institute collaborated with the UK government in 2012, advocating for an open data culture, developing standards and ethical frameworks for data openness, and achieving data sharing and circulation [13]. On the other hand, the digital rights management (DRM) takes technical protection measures for data to control the use, copying, and distribution of digital content, in order to prevent unauthorized behavior [14].

2.2. Data Protection Methods

Some researchers have studied the protection practices of data using existing IP laws or other laws. Cha pointed out that the challenges of protecting big data in relation to IP, reviewed the protection offered by IP and civil law, and noted the limitations of current laws [15]. Tsakalerou studied IP as a key driver of business performance [16]. Park proposed that by analyzing the limitations of data protection systems, the most relevant or helpful laws in Korean IP law should be found to protect data [17]. Tae proposed a plan to protect AI learning data through patent law or the Unfair Competition Prevention Act, and a plan to introduce a "data patent" application system in preparation for the era of big data [18]. Kim studied the necessity and possible practical directions of protecting big data in the existing copyright systems of Japan, the United States, and the European Union from the perspective of IP law [19].

These above existing studies have largely focused on examining the laws and regulations that various countries and regions have enacted to protect data-related rights and interests. Studies on the overall practical application and methods, directions, and strategies for engaging in rule-making activities of DIP remain limited. Given these limitations, further investigation is needed to address the following questions: (1) How do different countries operationalize data intellectual property protection in practice? (2) What are the limitations and challenges of current legal practices regarding data protection? (3) What are the key aspects that should be addressed regarding the application practices of data protection across different countries?

To address these questions, this study employs online research, literature review, and comparative analysis to examine the relevant laws and regulations on data empowerment issued by some countries and regions, focus on the applicability and limitations of laws related to DIP protection, and the current application practices and methods of DIP.

3. Research Objects and Methodology

Based on the existing researches mentioned above, this article aims at researching on the relevant laws and regulations issued and promulgated specifically for data empowerment and the applicability and limitations of laws related to data intellectual property protection, and current status and the practice methods of DIP by using the following methods:

(1) Literature research method: collecting, organizing, and analyzing data and information on DIP protection methods through literature review and database retrieval in Web of Science Core Collection, Google Scholar, Scopus, China National Knowledge Infrastructure (CNKI) using the keywords of "Big data", "Intellectual property", and "data property rights", etc., finding relevant research theories and practical cases, and sorting out the obtained target information.
(2) Network research method: using search engines of Google and Bing to find relevant information on DIP and visiting relevant websites to collect information on the policy documents, research reports, and international organization publications.
(3) Comparative analysis method: comparing the data information obtained, analyzing the characteristics, differences, advantages, and disadvantages

of DIP applying through different IP laws.

Search queries in the database mentioned above include: "data property right*" or "data ip right*" or "data intellectual property right*" (Topic); "Big data" or "data property rights" (Title) and protect* (Title); "Big data" or "data property rights" (Title) and "Intellectual property" (Topic).

This article conducted an initial screening of the retrieved results by reviewing titles and abstracts. Records were excluded if they were irrelevant to the research topic or object, had low relevance, were duplicates, lacked full-text availability.

4. Results & Discussion

4.1. Application Practices on Relevant Laws, Regulations about DIP Protection

4.1.1. Global DIP protection rules

The rapid advancement of AI and the exponential growth of data have prompted countries and regions around the world to introduce numerous policies and regulations on data protection. Some laws, regulations, and application practices related to data empowerment are shown in Table 1.

Table 1. Relevant laws, regulations, and application practices for data empowerment

Countries /regions	Laws and regulations	Application practice	Year
International treaties	WCT	Compilations of data or other material are protected, and the protection is without prejudice to any copyright subsisting	1996
	TRIPS	Compilations of data or other material which by reason of the selection or arrangement of their contents constitute intellectual creations shall be protected	1994
Europe	The Directive of the European Parliament and of the Council on the Legal Protection of Databases	For the first time, exclusive protection of data has been implemented in the form of legislative power, indicating that creative databases can be protected by copyright	1996
	Article 8 of the Charter of Fundamental Rights of the European Union	Determine the fundamental rights of personal data protection at the constitutional level	2000
	General Data Protection Regulation (GDPR)	Unless the data subject explicitly agrees, the use of sensitive data that can reveal an individual's race, political orientation	2018
	Data Governance Act	Promote the secure flow of data and realize the value of data	2020
	EU Data Strategy	Establish a shared space for public and industry data for data circulation and sharing	2020
	Proposal for a regulation on harmonized Rules on Fair Access and Use of Data	Promote a balance between data protection and other rights, and coordinate the relationship with existing IP protection	2022
Germany	Copyright Law	The neighboring rights in the Copyright Law protect databases that do not have originality	2013
	Directive on the Protection of Trade Secrets	Clarifying the definition, acquisition, restrictions of trade secrets, the situations on its use and disclosure, exceptions, etc.	2016

	Data space system	Create a shared and trusted space for data transactions, provide exchange channels for data transactions	2021
France	Annual Report of the French Council of State	Explicitly refuse to recognize personal data property rights	2014
Britain	Growing the Artificial Intelligence Industry in the UK	Proposed to establish a data investment governance framework through a data trust system for data circulation and sharing	2017
U.S.A	Access to Certain Court Records Electronically System (ACCESS) Act	A platform that has been designated as a "gatekeeper" to facilitate the flow and sharing of data	2021
China	Cybersecurity Law	Data classification, backup, encryption and assessment system of data have been specified	2017
	Article 27 of the Data Security Law	Data processing activities should be in accordance with laws and regulations, and establish data security management system	2021
	Twenty Measures on Data	Accelerate the construction of a data infrastructure system, propose a data property rights system	2022
Japan	Copyright Law	A database that is creative in the selection of information or system structure can be protected through copyright	2018
	Unfair Competition Prevention Law	Stipulates the improper acquisition, use, and disclosure of data	2018
India	Personal Data Protection Law	Classifying personal data, such as low-risk data and high-risk data	2019

The data protection policies in the countries and regions discussed above highlight the following two concerns.

(1) **Diverse rules for data circulation**. A globally unified governance framework for cross-border data flows is currently absent. Existing mechanisms are primarily embedded in bilateral or multilateral free trade agreements, forming a cooperative co-governance model. Some countries impose strict restrictions on the transfer of domestic sensitive data, while others prioritize privacy protection as the primary principle, strictly restricting cross-border data transfers.

(2) **The intersection of corporate data rights and competition law.** Monopolization and disorderly competition over data among digital platforms are hindering the efficient circulation of data as a factor of production. This intensifies the inherent tension between protecting corporate data rights and the need for data circulation.

4.1.2. Application Practice Methods of DIP

(1) **Compliance and governance of research data**. To address the dual imperatives of efficient circulation and security assurance in open sharing of scientific data, this study suggests establishing an institutional data repository. The repository should include a mechanism for identifying DIP rights in scientific data, whereby a rights-granting model can continuously incentivize creation and production while enhancing resource allocation efficiency.

(2) **Protection of AI training data, derivative data, and model-generated data**. For AI training data, copyright protection method may apply to works with originality. Data collections lacking originality may be protected under anti-unfair competition law if they are substantively and systematically extracted by competitors. Training data containing personal data is subject to personal information protection laws. For AI

derivative data, annotated data and cleaned datasets that possess originality may qualify as compilations protected under WCT/TRIPS. Technical information that is subject to adequate confidentiality measures is not generally known may be eligible for trade secret protection. For AI model-generated data, judicial practice tends to hold that AI-generated content does not constitute a work due to the absence of human authorship, unless a human has made substantive selections, modifications, or arrangements.

(3) **Globalization and standardization of DIP governance**. Countries and regions should strengthen exchanges and cooperation in the formulation of DIP protection rules, jointly participate in relevant rule-making, and promote the formation of a globally unified framework for DIP protection through bilateral or multilateral cooperation agreements.

4.2. Application and limitations of DIP practices in China

4.2.1. Limitations of DIP practices in China

Based on current judicial cases, the scope and limitations of some applicable laws on DIP protection in China are shown in Table 2.

Table 2: Applicable Legal Practices and Limitations of Data Intellectual Property Protection in China

Applicable Law		Scope of application	limitations
Civil Code		Treating data as an important aspect of personal information protection and privacy protection.	While data serves as privacy protection, its circulation and sharing are naturally restricted
Anti Unfair Competition Law		Mainly used in the business field to regulate the business behavior of competitors.	There is no competitive relationship and infringement behavior that cannot be applied, and there are also limitations that cannot promote data circulation.
Copyright Law	Copyright	Data compilation works that meet the requirements of originality (including databases and enterprise data)	The innovation of data cannot be guaranteed and is difficult to determine; The restrictions on rights stipulated by the Copyright Law also have limitations on data protection.
	neighboring right	The goal and value are consistent and the content is in line, meeting the criteria for determining the object of adjacent rights.	The determination of data types that include adjacency rights requires consideration of multiple factors for argumentation.
Trademark Law		Manage some data in accordance with relevant trademark registration laws.	Unable to distinguish data types belonging to the trademark category.
Trade secret		Data that meets the characteristics of trade secrets.	The circulation and sharing of data are restricted
Patent law		Can protect the use, application, and creation methods of data; Technical solutions involving data.	Not protecting the data itself; The constituent elements of patent protection have excessively high requirements for data.

China's current data protection policies and practices exhibit the following main characteristics.

(1) **Administrative penalties dominate data infringement cases.** Primarily guided by the statutory fine ranges set out in the Data Security Law and the Personal Information Protection Law (PIPL), enforcement actions mainly consist of orders to rectify and warnings, with only a limited number of cases involving substantial fines.

(2) **Efforts are increasingly shifting toward constructing a protection system tailored exclusively to data.** The current judicial protection practice in China is to integrate data into the existing IP legal framework system based on its type and attributes, and to protect data that meets applicable conditions through corresponding intellectual property laws.

4.2.2. Application Practice Methods of DIP in China

(1) **Establishment of DIP registration system**. Laws and regulations about DIP registration, certification, protection, value assessment, transaction, and pledge at the national level, along with certain local normative documents are enacted in China on data property registration in multiple provinces and cities designated as pilot regions. They provide clear guidance and norms for the collection, storage, use, and disclosure of data, ensuring the legality and security of data.

(2) **Cultivation of professional workforce for DIP services**. Universities are encouraged to offer interdisciplinary courses on DIP within law, information management, and related disciplines to cultivate interdisciplinary talents. Cooperation with IP industry associations could be conducted to launch a DIP service talent certification system. Furthermore, jointly build training bases with technology companies could be established to learn compliance review and value evaluation in real-world data scenarios.

(3) **Participation in DIP related activities for university libraries or IP information service centers**. The following such activities could be included: a) digital transformation of various types of data within their own university and library and surrounding enterprises, b) hierarchical classification of various types of data in order to select the appropriate data protection method, c) the data management and governance for the library and university, and d) the data literacy education for students, librarians and researchers in the university.

5. Conclusion and perspective

Establishing a data protection system is a significant innovation that faces contradiction, difficulties, and challenges between enhancing data circulation, sharing and protecting data rights. This study conducts an in-depth investigation of key laws, regulations, and application practices related to data empowerment in some countries or regions, and based on a discussion of various national approaches to data empowerment and protection, proposes focal points and methodological innovations to ensure the legality and legitimacy of data use and circulation. This study provides a systematic reference on current legal frameworks and application methods of DIP protection, highlighting practical gaps and limitations for future practice.

In the future, our wok could be extended to further aspects as follows:

(1) Construction of specialized DIP system. By clarifying the object, content, ownership, and infringement standards of DIP rights, the improvement of the DIP protection system and the development of the digital economy could be promoted in order to provide useful references for DIP.

(2) Case studies. By collecting practical cases of data protection that have been concluded, the limitation and solutions could be provided reference for future data protection dispute cases.

Data utilization is now evolving from traditional analytics and recommendations toward generation, decision-making, and autonomous action driven by large models. The improvement of data literacy ability is a link to enhance organizational management ability, data service ability, and education and training ability. Data literacy education policies and standard frameworks could be studied in the future from the institutional level.

References

[1] Dai SX. Data factor: main characteristics, driving effect and development path. Marxism & Reality. 2020(6):171-177. doi: 10.15894/j.cnki.cn11-3040/a.2020.06.032.

[2] Zhang JY, Bao Y. Thoughts on Intellectual Property Literacy Education Based on ACRL Higher Education Information Literacy Framework. Library World. 2022(5):61–67. doi: 10.14072/j.cnki.tsgj.2022.05.014.

[3] He PY, Yang L. The dilemma and countermeasures of enterprise data intellectual property protection in the digital economy era. Journal of Chongqing University of Technology (Social Science). 2023(6):80-90. doi: 10.3969/j.issn.1674-8425(s).2023.06.008.

[4] Feng XQ. Theoretical Interpretation and Construction of Data Property and Its Legal Regulation. Journal of Political Science and Law. 2021(04):81-97.

[5] Rae KS. Establishing the concept of data ownership in the data economy era and reviewing laws related to personal information protection. Law Review. 2022, 22(3):1-32. doi: 10.57057/LawReview.2022.09.22.3.1.

[6] Kim MA. A Study on the Legislation for Big Data Utilization in Korea. Business Law Review. 2022;36(4):69-94.

[7] Suo FT, Pan ZH. Certificate of Data Property Rights Taking Intellectual Property Rights as Reference. Journal of China University of Mining & Technology (Social Sciences). 2023,25(03):61-72. doi: 10.20089/j.cnki.issn.1009-105x.2023.03.006.

[8] Shan XG. The Chinese Option for Data Intellectual Property Rights. Frontiers. 2023(06):38-47. doi: 10.16619/j.cnki.rmltxsqy.2023.06.004.

[9] WIPO.WIPO Copyright Treaty (WCT) (1996). 1996, https://www.wipo.int/edocs/pubdocs/en/wipo_pub_226.pdf.

[10] World Trade Organization. Agreement on Trade-Related Aspects of Intellectual Property Rights (as amended on 23 January 2017). https://www.wto.org/english/docs_e/legal_e/31bis_trips_01_e.htm.

[11] Linera MAP, Meuwese A. Regulating AI from Europe: a joint analysis of the AI Act and the Framework Convention on AI. Theory and Practice of Legislation. 2025,13(3):292-311. doi: 10.1080/20508840.2025.2492524.

[12] Chen BH, Chen YL, Li JX. Regulatory Framework for Generative AI and Identification of Key Legal Risks in China. 2025. https://www.ctils.com/articles/18019.

[13] Open Data Institute. The ODI is a non-profit company committed to advancing trust in data across the spectrum - from closed to shared to open data. 2025. https://theodi.org/about-the-odi/.

[14] Feng X, Pan YJ, Xiao NA. The development and future of digital rights management: A review. Neurocomputing. 2025.658:131672. doi: 10.1016/j.neucom.2025.131672.

[15] Cha SY. A study on the protection of big data under intellectual property law. Korean Lawyers Association Journal. 2018,67(2):71-148. doi: 10.17007/klaj.2018.67.2.002.

[16] Tsakalerou M. Intellectual property as a key driver of business performance: the case of Singapore and its implications for innovation in a knowledge-based economy. International Journal of Learning and Intellectual Capital. 2018,15(1):83-92. doi: 10.1504/IJLIC.2018.088347.

[17] Park JS. The protect ability of new data such as big data and etc. from the perspective of intellectual property law. Journal of Industrial Property. 2019,58:77-129. doi: 10.36669/ip.2019.58.3.

[18] Tae JS. A Study on the Protection of AI Training Data - Focusing on Protection under the Intellectual Property Act. The Legal Studies Institute of Chosun University. 2021,28(1):65-102.

[19] Kim J. A Legal Considerations on the Protection of big-data under Property Law. Legal Theory & Practice Review. 2021,9(4):375-442.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261085

The Impact of Business-trip Duration on Turnover Intention of Knowledge Workers: A Mediation Analysis

Yong Li [a,1] and Guanghe Li[b]
[a] *Business School, Henan University, China*
[b] *Hongfujin Precision Electronics Co., Ltd., Zhengzhou, China*

Abstract. To clarify the mechanism through which business-trip duration affects turnover intention among knowledge workers in manufacturing enterprises, this study constructs a mediation model. With business-trip duration as the independent variable, turnover intention as the dependent variable, and work-family conflict as the mediating variable, this study explores the relational path among the above variables. The study population consists of 2038 knowledge workers employed by six manufacturing enterprises across five cities. A sample of 300 employees is selected using stratified sampling to complete the questionnaire survey, and 207 valid questionnaires are collected. One outlier is removed through multivariate outlier detection, leaving 206 valid questionnaires for subsequent model testing. The conclusions drawn are as follows: (1) Male knowledge workers have significantly longer business-trip duration than females. Among age groups, knowledge workers aged 31-40 show the highest turnover intention. In terms of education, knowledge workers holding graduate degrees or above demonstrate the highest turnover intention, whereas those with junior-college diplomas report the lowest. (2) There is a positive correlation between the duration of business trips and work-family conflict, a positive correlation between work-family conflict and turnover intention, and a positive correlation between the duration of business trips and turnover intention. (3) Work-family conflict plays a partial mediating role in the process where the duration of business trips affects turnover intention.

Keywords. Knowledge workers, business-trip duration, work-family conflict, turnover intention

1. Introduction

By 2025, China's manufacturing value added had surpassed 34.7 trillion yuan, accounting for approximately 25% of GDP. Among the employees in China's manufacturing industry, the proportion of those with a junior-college diploma or above increased from approximately 3.8% in 1995 to approximately 32% in 2025. While manufacturing enterprises have experienced rapid development, their high turnover rate has also garnered widespread attention in the industry. ZLJC has released research data indicating that in 2010, the average turnover rate among domestic enterprise employees was approximately 8.6%, while the turnover rate in manufacturing enterprises was as

[1] Corresponding Author: Yong Li, E-mail: excellentrichard@163.com

high as 23.4%. With the vigorous development of the knowledge economy, the quality of human capital profoundly influences the survival and development of enterprises, and the role of knowledge workers is particularly prominent. The loss of knowledge workers not only incurs direct recruitment and training expenses but also results in hidden losses such as the loss of experience, requiring enterprises to bear higher talent replacement costs. How to retain knowledge workers and effectively reduce their turnover rate has become one of the important issues facing enterprises. In the context of economic globalization, employees' work locations are no longer confined to a single region. The issue of employee turnover induced by business trips has gradually become prominent. For example, Guanghe Li provided a typical case: Company F is a well-known electronic manufacturing service enterprise headquartered in Shenzhen. Driven by business expansion, the headquarters transferred knowledge workers from its Zhengzhou subsidiary to Shenzhen for special project development in the form of business trips. The business travelers followed a "50+5" travel pattern: 50 days of travel, followed by 3-5 days of rest and recuperation in their home base before being dispatched again, repeating this cycle. The project lasted for 11 months. A department of the Zhengzhou subsidiary deployed 32 knowledge workers for this dispatch project. Twelve employees resigned upon project completion, yielding a turnover rate of 37.5%, which was substantially higher than the department's turnover rate prior to the project [1].

The aforementioned case reflects the correlation between business trips and employee turnover behavior. However, turnover behavior is an ex-post-facto outcome, which is not conducive to pre-emptive intervention by enterprises; turnover intention, as a key antecedent of turnover behavior, can reflect employee turnover risk earlier. Existing studies have shown that turnover intention is primarily influenced by multiple factors: individual factors (gender, age, education, personality traits, etc.), organizational factors (compensation and benefits, organizational commitment, etc.), individual-organization fit (job satisfaction), and external environment (employment opportunities, etc.). However, existing literature rarely explores the internal mechanism by which the duration of business trips affects turnover intention. In view of this, this paper sets the duration of business trips as the independent variable and turnover intention as the dependent variable to explore the relationship among the duration of business trips, work-family conflict, and turnover intention among knowledge workers in the manufacturing industry.

2. Concept Definition and Literature Review

2.1. Knowledge workers

Drawing on Peter Drucker's initial concept of knowledge workers and subsequent definitions of knowledge-based employees by scholars, this article defines knowledge workers as those who use knowledge to create value.

2.2. business-trip duration

This study combines the definitions provided in "Cihai" with existing relevant literature to define the duration of business trips as the cumulative time employees spend traveling to areas outside their permanent workplace to carry out work due to work

arrangements within a calendar year.

Li Xixi found through research that employees who devote too much time to work, such as overtime and business trips, which occupy the time they should spend with their families, will inevitably affect their relationships with their families and undermine the satisfaction of family life [2]. Lin Xiqing et al. discovered through a psychological health survey of employees at an airline company that irregular working hours, frequent overtime, and business trips led to poor communication among colleagues, causing various psychological barriers for employees [3]. Shi Zhan found through a survey of employees at an accounting firm that due to high intensity of overtime and frequent business trips, the attraction of excellent talents to the accounting firm has diminished [4]. Summarizing the above research findings, for employees who are on business trips, if the company fails to provide timely humanistic care and comprehensive welfare benefits, employees will gradually develop negative emotions and form a negative psychological experience.

2.3. Work-family conflict

The widely recognized definition of "work-family conflict" is provided by Greenhaus et al.: work-family conflict is a type of role conflict arising from the incompatibility of role pressures in the two domains of work and family [5].

Britt et al. found that uneven time allocation and negative emotions can affect the smooth transition between employees' work and family roles, leading to increased work pressure that may persist until employees leave work and enter their homes, ultimately resulting in work-family conflict [6]. Yang Tiantian et al. investigated the mediating role of work-family conflict among female nurses, and their research showed that turnover intention is influenced by work-family conflict. If hospitals or nurses take measures to balance work-family conflict, nurses' turnover intention will significantly decrease [7]. A review of the literature reveals that work-family conflict has negative impacts on employees. Firstly, work-family conflict can increase employees' role stress, which is detrimental to their physical and mental health. Secondly, employees' turnover intention is influenced by work-family conflict.

2.4. Turnover intention

Based on various literature, this study defines turnover intention as the psychological tendency of individuals to subjectively consider leaving an organization and giving up their current position.

Chen Min et al. found through research on various types of domestic enterprises (state-owned enterprises, foreign enterprises, joint ventures, etc.) that the more satisfied employees are with their subjective evaluation of their enterprise, the lower their turnover intention is, and vice versa [8]. Xu Yingyin found that when knowledge workers experience high levels of job burnout, it affects their job satisfaction, which in turn affects their turnover intention [9].

3. Theoretical Basis and Hypothesis Proposal

This study relies on the resource preservation theory and work-family boundary theory

to construct an analytical framework. The theoretical origins and the adaptation logic to the variables in this study are as follows.

3.1. Resource Preservation Theory

The resource preservation theory, proposed by Hobfoll, is a classic theory explaining the generation of individual stress, emotional exhaustion, and negative behavioral intentions. The core idea of this theory is that individuals have a stable motivation to actively acquire, maintain, and invest in various resources. When individuals face continuous depletion of resources, insufficient resource benefits, or failure to obtain positive returns from resource investment, they will experience psychological stress and negative cognition, which in turn can induce withdrawn behaviors. Individual resources include material resources, time, energy, emotional maintenance, and family relationships [10].

In the context of this study, if knowledge workers have long-term and high-frequency business trips, it will continuously consume their time and energy resources, squeeze family companionship and personal rest and rejuvenation space, resulting in sustained net resource depletion. When organizations cannot effectively compensate for resource losses through salary incentives, flexible systems, emotional support, and other means, employee stress accumulates, work experience deteriorates, and ultimately turnover intention increases. Therefore, the resource preservation theory provides the underlying theoretical logic for this study to explain the positive impact of business-trip duration on turnover intention.

3.2. Work-family boundary theory

The work-family boundary theory, proposed by Clark, is used to explain the mechanism of role switching and the logic of conflict generation between individual work domains and family domains. The theory states that work and family belong to two relatively independent social systems, with clear physical and psychological boundaries. Individuals achieve orderly transitions between work roles and family roles through stable boundary management. Once the boundaries are continuously permeated, broken, or blurred, individuals will experience role confusion and adaptation imbalance, ultimately leading to work-family conflict [11].

In this study, if knowledge workers engage in long-term and frequent business trips, the spatio-temporal boundaries between work and family will be disrupted, with work tasks continuously encroaching on private life domains. This leads to employees being unable to properly fulfill their family responsibilities and complete role transitions, resulting in a direct conflict between work demands and family needs. Based on this, the theory clearly explains the pathway through which the duration of business trips significantly increases the level of work-family conflict, providing a theoretical basis for the mediating mechanisms studied.

In summary, based on two classic theories, this study constructs a mediating transmission path of "business-trip duration - work-family conflict - turnover intention", explains the generation mechanism of turnover intention among knowledge workers, and proposes the following research hypotheses accordingly:

H1: The duration of business trips has a significant positive impact on the turnover intention of knowledge workers.

H2: There is a significant positive correlation between the duration of business

trips and work-family conflict.

H3: There is a significant positive correlation between work-family conflict and turnover intention.

H4: Work-family conflict partially mediates the relationship between the duration of business trips and turnover intention among knowledge workers.

4. Research Methodology

4.1. Research Samples and Data Collection

The study population consists of 2038 knowledge workers employed by six manufacturing enterprises across five cities: one in the Zhengzhou Airport Economy Zone, one in the Zhengzhou Processing Zone, one in Jiyuan, one in Hebi, one in Jincheng, and one in Taiyuan. According to Feng Xiaotian's discussion on sample size, for small-scale surveys, the sample size should be between 100 and 300 [12]. A total of 300 questionnaires were distributed in this survey, and the sampling method used was stratified sampling. The survey method combined online electronic questionnaires with offline paper questionnaires. Ultimately, 207 valid questionnaires were collected, with an effective response rate of 69%. Further employing the multivariate outlier detection methods, the outlier questionnaire with the number 114 was eliminated, and 206 questionnaires were finally included for statistical analysis. Among these knowledge workers who have entered statistical analysis, 60.2% are male and 39.8% are female. In terms of age composition, those aged 25-30 and 31-40 are the majority, accounting for 37.4% and 41.7% respectively. In terms of educational background, 88.3% have a college diploma or above.

4.2. Variable measurement

The work-family conflict scale, developed by Netemeyer et al., measures the impact of work on family from both time and stress perspectives, consisting of 8 items [13]. The turnover intention scale, developed by Mobley, includes 4 items [14]. Both scales are well-established and employ a Likert 5-point scoring system, with 1 indicating "strongly disagree" and 5 indicating "strongly agree".

4.3. Data analysis strategy

This study employed SPSS22.0, the process plugin, and Amos software to conduct a reliability and validity assessments, independent sample t-tests and one-way ANOVA based on demographic variables, Pearson correlation analysis, and mediation analysis on the questionnaire.

5. Empirical Analysis Results

5.1. Reliability and validity testing

The Cronbach's alpha coefficients for the two scales of turnover intention and

work-family conflict were 0.863, 0.855, respectively, indicating good reliability for the two scales.

A confirmatory factor analysis was conducted on the turnover intention scale. The factor loadings for the four items were 0.67, 0.84, 0.88, and 0.74, respectively, indicating that turnover intention could effectively explain the four observed items. The CR value was 0.866, and the AVE value was 0.619, indicating that the scale met the convergent validity criteria. The chi-square to degrees of freedom ratio was 2.438, the RMSEA value was approximately 0.08, the NFI was 0.988, the CFI was 0.993, the IFI was 0.993, and the TLI was 0.978, indicating that all fit indices met the criteria and the model shows acceptable fit.

A confirmatory factor analysis was conducted on the work-family conflict scale. The chi-square to degrees of freedom ratio was 2.263, the RMSEA value was 0.078, the NFI was 0.919, the CFI was 0.952, the IFI was 0.953, and the TLI was 0.933, indicating that all fit indices meet the standards and the model fits well. The factor loadings for the eight items are: 0.65, 0.65, 0.69, 0.63, 0.69, 0.70, 0.59, and 0.60. The convergent validity test results show that the composite reliability (CR) of work-family conflict is 0.855 and the average variance extracted (AVE) is 0.425. This construct includes eight measurement items: seven standardized factor loadings are greater than 0.6, while only one loading is 0.59, close to the ideal cutoff value of 0.6. Although the AVE of this construct is slightly below the ideal threshold of 0.5, its composite reliability reaches a satisfactory level. Following prior empirical practice, the convergent validity of work-family conflict is acceptable for subsequent model testing. Future research will focus on optimizing the Chinese wording of individual items during the localization process of the English scale to enhance the measurement quality of the scale.

5.2. Analysis of differences based on demographic variables

An independent sample t-test was conducted based on the gender of knowledge workers, and the test results are presented in Table 1.

Table 1. Independent sample T test for gender

Test variable	Male	Female	T	P
turnover intention	2.88±0.82	2.78±0.72	0.851	0.396
work-family conflict	3.30±0.63	3.18±0.55	1.441	0.151
business-trip duration	84±51 days	59±39 days	4.002	0.000**

Note: ** indicates P<0.01, and * indicates P<0.05

As can be seen from Table 1, there is a significant gender difference in business-trip duration, with male knowledge workers experiencing significantly longer time than females.

A one-way ANOVA was conducted based on knowledge workers' age, and the results are presented in Table 2.

Table 2. One-way analysis of variance for age

Test variable	25 years old and below	26-30 years old	31-40 years old	41-50 years old	F	P
turnover intention	2.64±0.65	2.90±0.79	2.95±0.79	2.25±0.58	3.463	0.009**
work-family conflict	3.23±0.67	3.26±0.59	3.26±0.61	3.23±0.53	0.036	0.997

business-trip duration	64±52 days	75±39 days	80±56 days	52±31 days	1.364	0.248

Note: ** indicates P<0.01, and * indicates P<0.05

As can be seen from Table 2, there is a significant age difference in knowledge workers' turnover intention. Among them, employees aged 31-40 have the highest turnover intention.

A one-way ANOVA analysis based on educational background was conducted, and the results are presented in Table 3.

Table 3. One-way analysis of variance for educational background

Test variable	High school and equivalent	Junior-college	Bachelor	Master and above	F	P
turnover intention	2.82±0.75	2.66±0.83	2.90±0.75	3.21±0.72	2.854	0.038*
work-family conflict	3.42±0.64	3.31±0.58	3.15±0.62	3.36±0.42	2.015	0.113
business-trip duration	65±39 days	79±49 days	72±50 days	70±47 days	0.656	0.580

Note: ** indicates P<0.01, and * indicates P<0.05

As can be seen from Table 3, there is a significant difference in turnover intention among knowledge workers based on their educational background. Employees with a graduate degree or above have the highest turnover intention, while those with a junior-college diploma have the lowest.

5.3. Correlation test

For "turnover intention", "work-family conflict", and "business-trip duration", a correlation analysis was conducted using the "Pearson correlation coefficient", and the significance test was performed using the "two-tailed test". The results are presented in Table 4.

Table 4. Correlation test

	turnover intention	work-family conflict	business-trip duration
turnover intention	1		
work-family conflict	0.293**	1	
business-trip duration	0.283**	0.333**	1

Note: ** indicates P<0.01, and * indicates P<0.05

From Table 4 above, it can be seen that there is a significant positive correlation between business-trip duration and work-family conflict, as well as between business-trip duration and turnover intention; there is also a significant positive correlation between work-family conflict and turnover intention.

5.4. Mediation Analysis

Using the Process Model 4 developed by Hayes, we examined the mediating effect of work-family conflict on the relationship between business-trip duration and turnover intention, while controlling for gender, age, and education. The results are presented in Table 5 and Table 6. The findings indicate that after incorporating the mediating variable of work-family conflict, the predictive effect of business-trip duration on

turnover intention is significant (t=2.82, p<0.01). Similarly, the predictive effect of business-trip duration on work-family conflict is significant (t=12.23, p<0.01), and the predictive effect of work-family conflict on turnover intention is also significant (t=3.42, p<0.01). Furthermore, business-trip duration can not only directly predict turnover intention but also predict it through the mediating effect of work-family conflict. The direct effect and mediating effect account for 70.37% and 29.63% of the total effect, respectively. The results provide support for Hypotheses H1-H4.

Table 5. Mediating effect test for work-family conflict

	Outcome turnover intention			Outcome turnover intention			Outcome work-family conflict		
Predictor	Coeff.	t	p	Coeff.	t	p	Coeff.	t	p
business-trip duration	0.19	2.82	0.005**	0.27	4.10	0.00**	0.26	12.23	0.00**
work-family conflict	0.30	3.42	0.001**						
R^2		0.23			0.18			0.15	
F		5.70**			4.77**			3.81**	

Note: ** indicates P<0.01, and * indicates P<0.05; The mediating model controls for gender, age, and education; among them, age and education are converted to dummy variables.

Table 6. The result of total effect, direct effect and mediating effect

	Effect	BootSE	BootLLCI	BootULCI	effect proportion
total effect	0.27	0.74	0.05	0.34	
direct effect	0.19	0.07	0.05	0.34	70.37%
The mediating effect of work-family conflict	0.08	0.03	0.03	0.14	29.63%

6. Conclusion and Discussion

6.1. Main research conclusions

This article focuses on knowledge workers who regularly undertake business trips to different locations, integrating multiple theories to construct a mediation effect model to examine the internal mechanism by which business-trip duration affects turnover intention. The main conclusions are as follows:

(1) There are significant differences in business-trip duration among knowledge workers based on gender, with male employees experiencing significantly longer business-trip duration than females. There are also significant differences in turnover intention among knowledge workers based on age, with those aged 31-40 exhibiting the highest level of turnover intention. Additionally, there are significant differences in turnover intention based on educational background, with knowledge workers with graduate degrees or above exhibiting the highest level of turnover intention, while those with a college degree exhibit the lowest level.

(2) The duration of business trips significantly and positively impacts the turnover intention of knowledge workers. The longer knowledge workers. accumulate in non-local business trips, the higher their turnover intention becomes. Continuous non-local business trips continuously deplete individuals' diverse resources and disrupt work-family balance, serving as a significant precursor to the loss of knowledge talent.

(3) Work-family conflict partially mediates the relationship between the duration of business trips and turnover intention. Excessively long business trips will break the boundary between work and family, giving rise to work interference and family conflict; and the conflict further increases the tendency to resign. In other words, the duration of business trips not only directly increases the tendency to resign, but also indirectly drives turnover by intensifying work-family conflicts.

6.2. Management Implications

(1) Optimize the business trip coordination mechanism and control the intensity of continuous remote assignments for knowledge workers

Enterprises should establish a business trip coordination ledger, rationally plan project progress, and avoid unlimited and uninterrupted long-term external assignments; implement a system of knowledge workers rotation for business trips, dynamically control the annual cumulative business trip time, reduce continuous resource depletion from the source, and mitigate the risk of excessive infiltration of work-family boundaries.

(2) Establish a work-family balance support system to alleviate the role conflict caused by geographic separation

For high-frequency business travelers, differentiated support policies will be implemented, including flexible compensatory leave, family visit guarantees, and remote flexible office mechanisms. Normalized humanistic communication channels will be established to promptly understand the family difficulties of business travelers, reduce the level of work interference and family conflicts, and weaken the internal driving force of turnover intentions.

6.3. Research limitations and future prospects

This study has some shortcomings, such as:

(1) Limitations of the survey sample. The distribution of the research sample in this study is limited, and it is expected that future research will be conducted in more regions and among more manufacturing enterprises.

(2) Limitations of the research model. This study did not consider additional influencing factors, such as the moderating effect of alternative job opportunities on turnover intention. It is anticipated that future research will refine the model to be more comprehensive and in-depth.

References

[1] Li Guanghe. The impact of business-trip duration on the turnover intention of knowledge workers in Q company[D]. Kaifeng: Henan University. 2023.
[2] Li Xixi. Flying together[D]. Shanghai: East China Normal University. 2014.
[3] Lin Xiqing, Gong Yifan. Research on the influencing factors and countermeasures of corporate employees' mental health[J]. China Personnel Science.2020. 26(02): 28-39.
[4] Shi Zhan. Research on employee motivation issues in enterprises[J]. Standard Science. 2019.12: 150-156.
[5] Greenhaus J H, Beutell N J. Sources of conflict between work and family roles[J]. Academy of Management Review. 1985. 10(1): 76-88.
[6] Britt T W, Dawson C R. Predicting work-family conflict from workload, job attitudes, group attributes, and health: a longitudinal study[J]. Military Psychology. 2005. 17(3): 203-227.

[7] Yang Tiantian, Li Xiaoyan, Yu Liping, et al. The mediating role of work-family conflict in the relationship between nurses' job stress and turnover intention[J]. Nursing Research. 2020. 34(03): 420-424.
[8] Chen Min, Zhang Yiwei, Su Wenyang, Jiang Weilian. The impact of core self-evaluation on turnover intention of clinical nurse specialists[J]. Chinese Journal of Modern Nursing. 2017. 23(07): 912-916.
[9] Xu Yingyin. The impact of job burnout, job satisfaction, and turnover intention on knowledge workers[D]. Xiamen: Xiamen University, 2025.
[10] Hobfoll SE. Conservation of resources: a new attempt at conceptualizing stress[J]. American Psychologist. 1989. 44(3): 513-524.
[11] Clark SC. Work/family border theory: a new theory of work/family balance[J]. Human Relations. 2000. 53(6): 747-770.
[12] Feng Xiaotian. Social research methods[M]. Beijing: China Renmin University Press; 2018: 126-150.
[13] Netemeyer RG, Boles JS, McMurrian R. Development and validation of work-family conflict and family-work conflict scales[J]. Journal of Applied Psychology, 1996, 81(4): 400-410.
[14] Mobley WH, Griffeth RW, Hand HH, Meglino BM. Review and conceptual analysis of the employee turnover process[J]. Psychological Bulletin. 1979. 86(3): 493-522.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261086

A Novel Supply Chain Packaging Design Evaluation Method via Sentiment Analysis of Reviews

Yangqing LIU [a], Haifeng SONG [a], Xiumeng LI [a], Yaxi MA [a], Jingqing DU [a], Shenghai KE [a] and Weiwei YANG [b,1]

[a] *School of Art and Design, Taizhou University, Taizhou, 318000, Zhejiang, China*
[b] *School of Artificial Intelligence, Taizhou University, Taizhou, 318000, Zhejiang, China*

Abstract. This study uses online reviews to evaluate e-commerce packaging design. Amazon Reviews'23 records from the All_Beauty category are cleaned, screened with a packaging vocabulary list, and classified by a BERT sentiment model. The model reaches 91.3% accuracy and a 0.92 F1-score. Negative packaging reviews are then used to calculate failure rates, identify frequent complaint terms, and mine association rules with Apriori. The results point to repeated problems such as crushed boxes, leakage, seal or cap failure, wet cardboard, and soiled products. The method is not a replacement for physical tests. It is a way to find packaging failures that customers have already reported.

Keywords. Supply Chain Management, Package Evaluation, Sentiment Analysis, BERT, Apriori

1. Introduction

Packaging protects products during storage, transportation, and use [1]. Existing studies assess packaging performance through storage experiments [2] and field testing [3]. These methods are necessary, but they are costly and usually work under controlled conditions. In e-commerce, some failures only become visible after products pass through warehouses, carriers, and local delivery. Reviews then become a useful source of detail: users describe leakage, broken caps, crushed boxes, poor seals, wet cardboard, and delivery damage in plain language.

This study follows earlier review-based packaging research [4], but it does not rely only on word frequency. It combines Pack-List filtering, BERT sentiment analysis, failure-rate calculation, and Apriori association rules. This study focuses on three specific objectives: identifying packaging-related reviews, classifying their sentiment polarity, and mining failure term co-occurrences. Rather than providing a definitive engineering diagnosis, this method generates a prioritized list of potential failures to guide subsequent empirical testing.

The workflow is deliberately staged because each step reduces a different kind of noise. The original review corpus is large, only part of it concerns packaging, only

[1] Corresponding Author: Weiwei Yang.

part of the packaging subset is negative, and only some negative terms form useful co-occurrence patterns. This sequence keeps the method from becoming a general review classifier and makes it closer to a screening tool for packaging diagnosis.

2. Literature Review

The related work falls into three areas. The first is packaging evaluation. Impact simulation can reduce part of the test cost, although the result still depends on model assumptions [5]. Mechanical-property analysis helps connect packaging materials with handling conditions [6]. Environmental assessment tools further support packaging development decisions [7]. The second area is sentiment analysis. E-commerce sentiment studies show that deep models can capture review polarity at scale [8]. Earlier methods such as Naive Bayes are simple and interpretable [9]. CNN and CNN-LSTM models improve text representation but still require careful domain adaptation [10,11]. Multi-feature deep models add another way to combine lexical and semantic signals [12]. Transformer and transfer-learning models provide stronger contextual modeling [13,14]. Short-text noise and domain mismatch remain difficult in sentiment analysis [15].

The third area is e-commerce review research. Previous studies have discussed why consumers use online reviews [16]. Other work examines review helpfulness, images, and reviewer expertise [17]. Review sentiment has also been used for sales performance prediction [18]. Most of this work studies purchase behavior or marketing outcomes. Packaging failure is usually not the main topic. This study uses that review-mining logic for a more specific purpose: packaging failure diagnosis.

3. Methodology

Figure 1 shows the workflow. Raw reviews are cleaned by converting abbreviations, standardizing lowercase text, removing non-letter characters, correcting spelling, deleting stop words, and tokenizing the text. A Pack-List is then built from low-rated reviews and packaging terms such as box, wrapping, shipping, damaged, leak, broken, wet, crushed, cardboard, seal, and cap. This filter keeps the analysis away from comments about fragrance, price, color, or product performance.

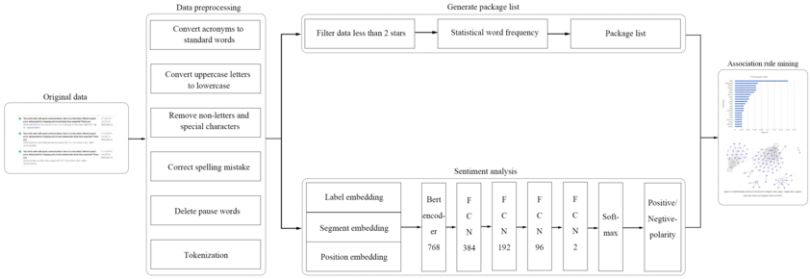


Figure 1. The architecture of the proposed method.

BERT is used for sentiment classification because packaging complaints depend on context [13,14]. Unlike simple frequency-based methods, BERT effectively captures

contextual negations (e.g., 'not damaged') and domain-specific nuances, such as distinguishing 'arrived leaking' from vague complaints about product quality. As shown in Figure 2, each review is converted into token, segment, and position embeddings before entering the encoder. The final hidden state of the [CLS] token is passed to a fully connected layer with sigmoid activation. The model is fine-tuned with AdamW, a 2e-5 learning rate, batch size 32, three epochs, and a 10% warmup schedule. Given the imbalanced nature of the dataset, where negative reviews dominate the filtered subset, the macro-averaged F1-score is emphasized to evaluate the model's robustness across both minority and majority classes.

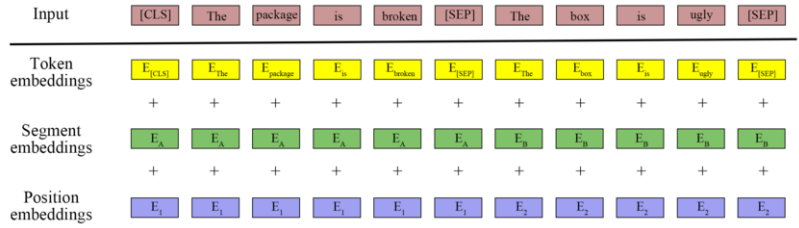


Figure 2. BERT input representation.

After classification, the negative packaging reviews are mined with Apriori. To prepare the data for association rule mining, each negative polarity review is treated as a transaction. We utilize the constructed Pack-List to filter the review text, transforming each negative review into an itemset. A rule $X \rightarrow Y$ indicates a directional association where the presence of antecedent X implies the likely presence of consequent Y in the review set. Support measures how often the itemset appears. Confidence measures how often Y appears when X appears. Rather than providing a definitive engineering diagnosis, this method generates a prioritized list of potential failures to guide subsequent empirical testing.

The rule interpretation also separates observed failure phenomena from possible packaging or logistics conditions. Words such as crushed, damaged, leaking, wet, and broken describe what users report seeing. Words such as shipped, box, cardboard, cap, and seal point toward where the failure may have started. The distinction is not perfect because reviews are informal, but it helps the analysis move from counting complaints to forming testable design hypotheses.

4. Empirical Analysis and Results

The empirical case uses Amazon Reviews'23. The All_Beauty subset contains more than 700,000 records, including rating, title, text, image, asin, user_id, timestamp, verified_purchase, and helpful_vote fields. After packaging-related reviews are filtered, the data are split into training and testing sets at an 8:2 ratio. We compared our proposed model (BERT with Self-Attention) with the following baselines: SVM using TF-IDF features, LSTM for sequence processing, and the standard pre-trained BERT-base model. Table 1 shows that the proposed model performs better, reaching 91.3% accuracy and a 0.92 F1-score, demonstrating that the integration of the self-attention mechanism effectively enhances the model's ability to capture semantic features.

Table 1. Performance comparison of different models.

Model	Accuracy	Precision	Recall	F1-score
SVM	85.2%	0.84	0.85	0.84
LSTM	89.5%	0.88	0.90	0.89
BERT-base	90.1%	0.91	0.92	0.91
Proposed Model	91.3%	0.92	0.93	0.92

Figure 3 summarizes the Pack-List vocabulary. The main terms are linked to damaged boxes, leakage, broken caps, crushed packages, wet cardboard, poor seals, and shipping damage. Figure 4 shows the sentiment distribution. In the filtered packaging-related subset, 82.68% of reviews are negative and 17.32% are positive.



Figure 3. Word cloud diagram of the packaging list.

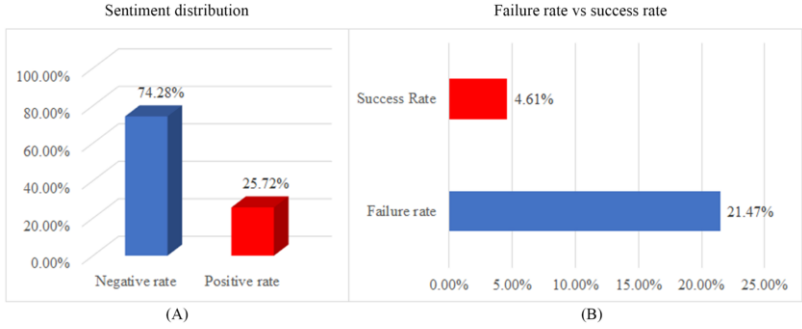


Figure 4. (A) Sentiment distribution over filtered packaging-related reviews; (B) packaging failure and success rates.

The monthly packaging failure rate is calculated as negative packaging reviews divided by total reviews in the same month. Figure 5 shows peaks in August 2020, October 2020, and January 2022. These peaks may be related to logistics pressure, material changes, storage problems, or transport disruption. The time pattern helps connect review complaints with possible changes in logistics, inventory, weather, or packaging batches.

Negative word-frequency analysis narrows the failure vocabulary. Figure 6 shows that damaged, broken, leaking, crushed, wet, and box-related terms appear repeatedly

in negative packaging reviews. These words connect customer complaints with design issues such as compression strength, sealing performance, moisture protection, and handling damage.

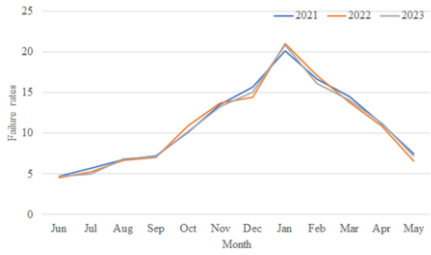


Figure 5. The trend of packaging failure rate over time.

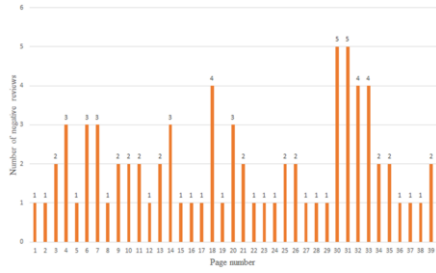


Figure 6. Negative polarity word frequency analysis.

For association rule mining, the minimum support is set to 0.02 and the minimum confidence to 60%. These thresholds were identified through iterative analysis. The relatively low support threshold ensures that infrequent but critical failure modes, such as specific leakage issues, are not ignored in the massive dataset, while the 60% confidence maintains strong predictive validity. Figure 7 and Table 2 summarize the main rules. {Crushed, Shipped} → {Heavily Damaged} suggests weak compression or stacking resistance. {Box, Leak} → {Cap or Seal Broken} points to closure-integrity problems. {Cardboard, Wet} → {Soiled} suggests moisture exposure during transport.

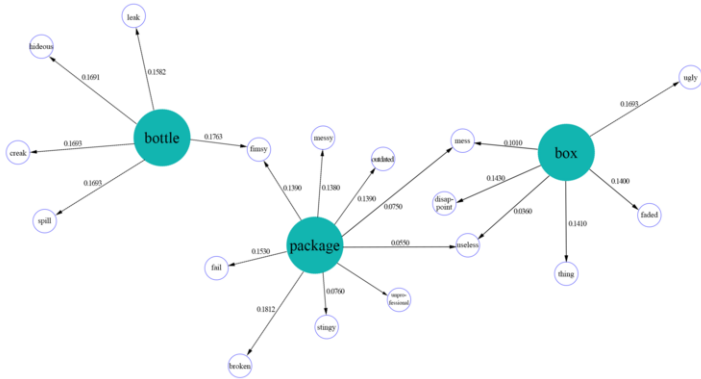


Figure 7. Association strength analysis.

Table 2. High-impact association rules derived from negative reviews.

Antecedent (X)	Consequent (Y)	Support (%)	Confidence (%)	Lift
{Crushed, Shipped}	{Heavily Damaged}	0.08	85.4	4.2
{Box, Leak}	{Cap or Seal Broken}	0.05	72.1	3.5
{Cardboard, Wet}	{Soiled}	0.04	65.8	2.8

The results should be read as a sequence rather than as separate outputs. The model comparison supports the sentiment labels used downstream. The sentiment distribu-

tion explains why negative packaging reviews are the focus of the failure analysis. The time-series curve identifies months in which complaints become unusually dense. Word-frequency analysis then shows which failure terms dominate, and association rules show which terms form stable combinations. Together, these steps turn an unstructured review corpus into a smaller set of packaging issues that can be checked against materials, sealing design, warehouse conditions, and carrier handling.

5. Discussion and Conclusion

The method is a complement to physical packaging tests. Drop tests, compression tests, material tests, and simulation are still needed, but review mining shows what users report after products pass through warehouses, carriers, seasonal peaks, and local handling. For designers, the rules give a starting point for checking carton strength, cushioning, cap torque, seal quality, moisture resistance, and shipping protection.

For supply chain management, the practical use is straightforward. A company can monitor packaging-related negative reviews by month, compare peaks with logistics records, and then decide whether to inspect a supplier, a shipping route, a warehouse condition, or a packaging specification. The method does not claim that a package failed because of one single cause. It gives a ranked set of likely problems, which is enough to guide targeted testing more efficiently than testing every possible packaging scenario without knowing where customers actually saw failures.

The study has limits. The All_Beauty category may not represent food, fragile goods, heavy products, or large liquid containers. The data are also shaped by language and review bias, since dissatisfied users are more likely to write. Future work should test multilingual models, build category-specific Pack-Lists, and validate the method on more platforms. The next step is to compare the mined rules with physical tests, rather than treating review patterns as final evidence.

Funding

This research was supported by the National Social Science Foundation Art General Project of China (23BG123), Zhejiang Provincial Department of Science and Technology Soft Science Research Program (2025C35056), the Science and Technology Planning Projects of Taizhou (24nya03), Zhejiang Provincial Educational Science Planning Project (2025SCG159), the Innovation and Entrepreneurship Training Program (2024103500046), Zhejiang Province Philosophy and Social Sciences Planning Project (26NDJC206YB), the Education and Science Planning Project of Taizhou (TG24032), the second round of provincial teaching reform projects for undergraduate and graduate education under the 14th Five-Year Plan (JGBA2024522), and the Ministry of Education Industry-Academia Cooperative Education Program (230824464907176).

References

[1] Awol SM, Kuyu CG, Bereka TY, Abamecha N. Physicochemical stability, microbial growth and sensory quality of refined wheat flour as affected by packaging materials during storage. Journal of Stored Products Research. 2024;105:102217, doi:10.1016/j.jspr.2023.102217.

[2] Kumar MN, Jain R, Singh M, Tiwari A, Singh B, Sethi S, Sathee L, Lal K, Kumar A, Madhavi K. Evaluation of packaging materials for enhancing the storage life of marigold flowers (Tagetes erecta). The Indian Journal of Agricultural Sciences. 2024;94(2):192-197, doi:10.56093/ijas.v94i2.142634.
[3] Faye SLB, Lugand MM, Offianan AT, Dossou-Yovo A, Kouadio DKM, Pinto F. Field testing of user-friendly perennial malaria chemoprevention packaging in Benin, Côte d'Ivoire and Mozambique. Malaria Journal. 2024;23(1):157, doi:10.1186/s12936-024-04977-0.
[4] Esfahanian S, Lee E. A novel packaging evaluation method using sentiment analysis of customer reviews. Packaging Technology and Science. 2022;35(12):903-911, doi:10.1002/pts.2686.
[5] Ballo F, Previati G, Mastinu G, Comolli F. Impact tests of wheels of road vehicles: A comprehensive method for numerical simulation. International Journal of Impact Engineering. 2020;146:103719, doi:10.1016/j.ijimpeng.2020.103719.
[6] Fadiji T, Berry T, Coetzee CJ, Opara U. Investigating the mechanical properties of paperboard packaging material for handling fresh produce under different environmental conditions: Experimental analysis and finite element modelling. The Journal of Applied Packaging Research. 2017;9:20-34.
[7] Molina-Besch K, Pålsson H. A simplified environmental evaluation tool for food packaging to support decision-making in packaging development. Packaging Technology and Science. 2020;33(4-5):141-157, doi:10.1002/pts.2484.
[8] Chamekh A, Mahfoudh M, Forestier G. Sentiment analysis based on deep learning in e-commerce. Springer-Verlag; 2022. p. 498-507, doi:10.1007/978-3-031-10986-7_40.
[9] Saepudin S, Widiastuti S, Irawan C. Sentiment analysis of social media platform reviews using the Naïve Bayes classifier algorithm. Jurnal Sisfokom (Sistem Informasi dan Komputer). 2023;12(2):236-243, doi:10.32736/sisfokom.v12i2.1650.
[10] Rani S, Kumar P. Deep learning based sentiment analysis using convolution neural network. Arabian Journal for Science and Engineering. 2019;44(4):3305-3314, doi:10.1007/s13369-018-3500-z.
[11] Ma X. A novel hybrid model by using convolutional neural network and long short-term memory for text sentiment analysis. Dyna. 2020;95(1):527-533, doi:10.6036/9714.
[12] Abdi A, Shamsuddin SM, Hasan S, Piran J. Deep learning-based sentiment classification of evaluative text based on multi-feature fusion. Information Processing and Management. 2019;56(4):1245-1259, doi:10.1016/j.ipm.2019.02.018.
[13] Brauwers G, Frasincar F. A survey on aspect-based sentiment classification. ACM Computing Surveys. 2022;55(4):Article 65, doi:10.1145/3503044.
[14] Hartmann J, Heitmann M, Siebert C, Schamp C. More than a feeling: Accuracy and application of sentiment analysis. International Journal of Research in Marketing. 2022;40, doi:10.1016/j.ijresmar.2022.05.005.
[15] Poria S, Hazarika D, Majumder N, Mihalcea R. Beneath the tip of the iceberg: Current challenges and new directions in sentiment analysis research. IEEE Transactions on Affective Computing. 2023;14(1):108-132, doi:10.1109/TAFFC.2020.3038167.
[16] Changchit C, Klaus T, Lonkani R. Online reviews: What drives consumers to use them. Journal of Computer Information Systems. 2020;62:1-10, doi:10.1080/08874417.2020.1779149.
[17] Bigne E, Ruiz C, Cuenca A, Perez C, Garcia A. What drives the helpfulness of online reviews? A deep learning study of sentiment analysis, pictorial content and reviewer expertise for mature destinations. Journal of Destination Marketing and Management. 2021;20:100570, doi:10.1016/j.jdmm.2021.100570.
[18] Yuan H, Xu W, Li Q, Lau R. Topic sentiment mining for sales performance prediction in e-commerce. Annals of Operations Research. 2018;270, doi:10.1007/s10479-017-2421-7.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)

© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261087

AI-Empowered Value Co-Creation for Intangible Cultural Heritage Digital Revitalization: A Conceptual Multi-Case Analysis

Mingjia Lin[a,1] , Jianhong Sun[a], Chunqiong Wu[a] and Tianhua Hui[a]
[a] *School of Economics and Management, Yango University, Fuzhou, China*
ORCiD ID: Mingjia Lin https://orcid.org/0009-0001-5497-1246
Chunqiong Wu https://orcid.org/0009-0003-8621-9029

Abstract. The digital revitalization of intangible cultural heritage (ICH) is central to China's national strategy for digital culture. Artificial intelligence (AI) is reshaping how ICH is safeguarded, transmitted, and innovated. Yet most studies focus on technological applications and have not systematically theorized how AI enters ICH value co-creation. Drawing on value co-creation theory, this study develops an analytical framework. It applies the framework through a conceptual multi-case analysis of eight publicly accessible online cases spanning different ICH domains and lead actors. Cross-case comparison tests the framework's explanatory power. The findings identify four dominant models of ICH value co-creation under AI: heritage-bearer-led human-machine co-creation, public-platform-based crowd co-creation and benefit sharing, business-driven value monetization, and government-led ecosystem synergy. Multiple actors advance ICH digital revitalization through resource co-creation, production co-creation, experience co-creation, and value sharing, and AI shifts from an instrumental tool to a collaborative partner. The study extends value co-creation theory and offers an analytical tool for ICH digital revitalization.

Keywords. ICH digitalization; value co-creation; artificial intelligence

1. Introduction

ICH is a vital part of China's traditional culture, and its cultural genes enhance cultural confidence and soft power. In setting the strategic goal of building a cultural power, the Chinese leadership has called for "exploring effective mechanisms for integrating culture and technology" [1]. The aim is to empower cultural development digitally, transform it through informatization, and convert cultural resource advantages into cultural development advantages. UNESCO has likewise affirmed the fundamental role of digital

[1] Corresponding Author: Mingjia Lin, 398479760@qq.com.
This work was supported by (1) the 2025 Fujian Provincial Education Research Project for Young and Middle-Aged Teachers, “Research on the Construction and Application of an Intangible Cultural Heritage Lacquer Art Knowledge Graph Based on RAG Optimization” (Project No. JAT251231); and (2) the 2024 Fujian Provincial Science and Technology Plan Project, “Research on the Transformation Mechanism and Pathways of the Digitization of the Agricultural Industry Chain Driven by New E-Commerce” (Project No. 2024R0071).

means in heritage safeguarding and transmission [2]. Two instruments make this clear: the Convention for the Safeguarding of the Intangible Cultural Heritage and the Charter on the Preservation of Digital Heritage. AI is moving ICH from digital storage toward intelligent innovation, making digital revitalization a core task for the field.

Existing studies of ICH digitalization concentrate mainly on technological applications. Value co-creation theory holds that ICH value emerges not from a single actor but through interaction among multiple actors and the integration of resources. This perspective provides a useful lens for ICH digital revitalization. Yet few studies ask how AI, with its autonomy, learning capacity, and generativity, becomes embedded in and reshapes value co-creation among ICH actors. This study therefore addresses two questions. First, what dominant models of value co-creation emerge when AI enters ICH digital revitalization? Second, through which core mechanisms do multiple actors drive ICH digital revitalization and share value?

To answer these questions, the study builds an analytical framework on value co-creation theory and tests it with eight publicly accessible online cases. The contribution is threefold. It extends value co-creation theory into AI-empowered ICH digital revitalization and clarifies AI's role in multi-actor interaction. It identifies four dominant models and four core mechanisms that offer an operational basis for future empirical research. And it grounds the framework in systematic cross-case comparison.

2. Theoretical Foundations and Literature Review

2.1. ICH digital revitalization: connotations and features

The digitalization of ICH uses capture, storage, processing, display, and dissemination technologies to convert, represent, and restore ICH in shareable, reproducible digital forms. Digital safeguarding must respect the living nature of ICH [3]. As safeguarding has moved from static rescue preservation to dynamic productive preservation, ICH digitalization has evolved from digital archiving to digital presentation and digital experience [4]. This trajectory aligns with digital transformation theory, which treats technology, organization, and culture as changing together [5]. Digital revitalization goes further: it restores the authenticity and original character of ICH, turns static cultural archives into callable, shareable, reproducible digital assets, and realizes the creative transformation and innovative development of traditional culture. Craft-based ICH is rich in embodied knowledge, and preserving its authenticity through digitization remains a shared challenge in international digital heritage research [6].

Digital revitalization proceeds through three progressive layers: resourceization, mediatization, and vitalization. Resourceization uses databases and digital archives to build classification standards and retrieval systems for ICH [7]. Mediatization uses networks, mobile devices, and public platforms to expand ICH's expressive space through interactivity, convenience, and immersion [8]. Vitalization embeds digital technology in the existence and development of ICH [3], keeping it alive today. Digitalization thus does not lock ICH away; it builds a cultural ecosystem that multiple actors can access and co-create. This is why the value co-creation perspective is needed.

2.2. Value co-creation theory and its application to ICH

Value co-creation theory has evolved from customer-experience-based co-creation [9] and service-dominant-logic-based co-creation [10] to service-ecosystem-based co-creation [11]. The unit of analysis has expanded from dyadic firm-customer interaction to multi-actor networks. A service ecosystem is a loosely coupled structure in which actors co-create value through institutions, technology, and language on the basis of value propositions [8]; later work brought institutions into this view [12]. ICH research has already examined multi-actor logics and co-creation mechanisms. Wen and Zhao (2022) [4] show that actors differ in their advantages in data, technology, industrial application, and talent. Song et al. (2022) [13] map the roles of government, scenic areas, heritage bearers, and tourists and highlight the asymmetry of value co-creation. Ma and Chen (2025) [14] find that multiple actors jointly perform value co-creation actions.

These studies move ICH value co-creation from single-actor analysis toward a systematic view of multi-actor dynamics, but two gaps remain. First, most studies examine digital forms such as short videos, e-commerce livestreaming, and digital museums; few theorize how AI enters the co-creation process, changes actor interaction, or improves resource integration. International research, by contrast, has discussed AI's transformative impact [15], the changing role of AI in services [16], and the need for context-sensitive frameworks integrating AI with value co-creation [17]. Second, the actor that uses technology has been underplayed. Jian et al. (2016) [18] treat technology as a dynamic operant resource that promotes interaction and value co-creation at multiple levels. AI, however, is autonomous, generative, and capable of learning; it is no longer merely a tool to be operated but an important nexus of value co-creation. Existing frameworks do not fully explain this shift. This study therefore adopts the service ecosystem perspective, treats AI as an empowering element, and builds an analytical framework of value co-creation for AI-empowered ICH digital revitalization.

3. Research design

This study adopts a conceptual multi-case analysis. Rather than generating new theory from data, this approach applies an existing theoretical perspective to a new context and uses typical cases to enrich the explanatory power of a conceptual framework. The conclusions take the form of a conceptual framework with illustrative evidence.

Case selection followed four criteria: information accessibility, domain diversity, actor diversity, and technological diversity. Eight publicly accessible online cases span traditional craftsmanship, traditional fine arts, traditional drama, and folk customs, involving heritage bearers, public platforms, business actors, and government, and technologies including digital twins and generative AI. Data from government sites, heritage bodies, media, and corporate releases were cross-verified. All cases are Chinese, so generalizability awaits cross-national testing.

The analysis proceeded in three steps. First, each case was organized into an evidence chain covering actor roles, AI involvement, co-creation behaviors, and value outcomes; second, themes such as actor roles, AI technologies, and co-creation mechanisms were identified; third, cross-case comparison tested and refined the conceptual framework. A case-evidence alignment (Table 1) ensured the traceability and credibility of the conclusions.

4. Findings

4.1. Four dominant models

Cross-case comparison shows that the eight cases share a common logic despite differences in domain, technology, and context (Table 1). In each case, a core actor leads value co-creation, and AI enters the resource, production, and experience stages under that actor's coordination, producing identifiable value outcomes. On this basis, four dominant models can be distinguished.

Table 1. Cross-case comparison.

Case	Dominant model	AI empowerment approach	Main mechanisms	Value outcomes
West Lake Longjing tea	Human-machine co-creation	Digital modeling of craft	Resource co-creation; production co-creation	Tacit skills made explicit; digital archive created
Quanzhou Jishiban	Human-machine co-creation	Generative content creation	Experience co-creation	Digital dissemination of the custom; broad public participation
Houma shadow puppetry	Crowd co-creation and benefit sharing	Interactive display	Experience co-creation; resource co-creation	ICH shifts from being viewed to being co-created
AI thangka restoration	Crowd co-creation and benefit sharing	Intelligent restoration	Resource co-creation; production co-creation	Higher restoration efficiency; public participation in safeguarding
Shu brocade and Shu embroidery	Value monetization	Intelligent design; flexible production	Production co-creation; value sharing	Personalized products; improved market performance
Baidu Baike AI ICH Pavilion	Value monetization	AI agents; digital human	Experience co-creation; value sharing	Wide knowledge dissemination; tailored experiences
Shibing Miao embroidery	Ecosystem synergy	Image recognition; generative algorithms	Resource co-creation; production co-creation; value sharing	Industrial agglomeration; employment and income gains
Quanzhou string puppetry	Multiple mechanisms combined	Knowledge base; intelligent evaluation	Resource co-creation; experience co-creation	Open resource sharing; public learning and practice

4.1.1. Heritage-bearer-led "human-machine co-creation"

This model centers on the heritage bearer. Heritage bearers hold the craft knowledge and cultural judgment; AI performs technical tasks such as data capture, modeling, and generation. The two form a collaboration in which the bearer decides, and AI executes and feeds back. In the West Lake Longjing tea case, the ten key hand movements depend heavily on individual experience. Heritage bearer Tang Xiaojun worked with a technology firm to capture hand trajectories, force changes, and fire control through AI super-assembly algorithms and high-precision sensing. Three-dimensional modeling and digital twins turned these data into a quantifiable, demonstrable Longjing Digital Archive. Hard-to-verbalize sensory experience thus became recordable, transmissible digital models, and tacit knowledge was externalized into shared public knowledge. In the Quanzhou Jishiban case, a ritual custom transmitted orally, provincial-level bearer Guo Qinghai worked with a radio and television station to produce an AI microfilm. The film

drew on the custom's historical origins, folk connotations, and spiritual core and promoted its digital dissemination.The two cases share a common logic: the bearer keeps control of cultural meaning and craft standards, while AI's retrieval and generation capabilities broaden the bearer's cognitive horizon and speed skill development. The cases differ in that craft-based ICH relies more on sensing and modeling, whereas custom-based ICH relies more on generative content. Human-machine co-creation thus presupposes that bearers retain cultural decision-making authority, so that technology does not alienate ICH authenticity.

4.1.2. Public-platform-based "crowd co-creation and benefit sharing"

This model centers on public platforms such as heritage institutions, schools, and communities. These platforms build ICH databases and creation tools and open AI capabilities to the public, allowing non-professionals to re-create and re-disseminate ICH content. Anyone can participate, and the resulting value is shared by the public. In the Houma shadow puppetry case, AIGC models generated an interactive display that augmented reality (AR) rendered dynamically. Visitors could interact with digital characters and scenes in a small space, and ICH shifted from being viewed to being co-created. In the thangka restoration case, a technical team built an AI restoration system on Baidu's Paddle platform. The system generated restoration references from historical thangka data, letting non-experts join restoration work and markedly improving its efficiency. Both cases show AI empowering non-experts: the threshold of participation drops sharply, and the public changes from cultural recipients into co-creators. The model's effectiveness depends on platform affordances, and the depth of participation on platform openness and public digital literacy.

4.1.3. Business-driven "value monetization"

This model centers on enterprises. Driven by market logic, firms integrate ICH craftsmanship with modern technology, capture economic returns while safeguarding ICH, and create a sustainable commercial cycle. In the Shu brocade and Shu embroidery case, Sichuan Shujing Cultural Communication Co., Ltd. used AI to enrich its pattern library, generate design drafts, and run small-batch flexible production, shortening the production cycle and enabling personalized customization that brought Shu embroidery closer to mass demand. In the Baidu Baike AI ICH Pavilion case, a commercial platform and cultural and tourism departments launched AI agents for Peking opera, the Xunpu flower headdress, Chinese chess, and other cultural forms. They also released a digital human of a national-level Bai tie-dye bearer. The agents offer knowledge Q&A, creative interaction, and craft demonstration, presenting a new form of AI-ICH experience. Both cases show firms using brand-based operation to fill the funding, talent, and channel gaps of traditional safeguarding models and using AI to build emotional ties between ICH and consumers. Market logic can nevertheless distort ICH's cultural meaning, and excessive commercialization must not be allowed to erode its authenticity.

4.1.4. Government-led "ecosystem synergy"

This model centers on government. Through policy, institutional design, and funding, government builds an institutional ecosystem that aligns bearers, enterprises, universities, and the market. It is the most systemically integrated of the four models. With policy support, Shibing County in Guizhou built the world's first digital gene bank for

safeguarding Miao embroidery. The bank collected patterns from more than a thousand bearers. It applied algorithms such as Retinex-Net and GAN so that AI could grasp the cultural meaning and aesthetic features of Miao embroidery. The county also built an ethnic embroidery industrial park and the Wisdom Brocade: Guizhou Women's Characteristic Handicraft Innovation and R&D Base, forming a complete Miao embroidery plus tourism chain. Under this model, government supplies public resources and operating rules, and dispersed actors perform their roles and coordinate within the institutional ecosystem, enabling value co-creation at scale. The model typically requires an industrial base and strong governance capacity, and it works in combination with resource co-creation, production co-creation, and value sharing.

4.2. Core mechanisms of value co-creation in ICH digital revitalization

One case often activates several mechanisms at once (Table 1). Resource co-creation is the foundation on which value co-creation happens; production co-creation is where value is generated; experience co-creation is where value circulates; and value sharing is where value is realized. Together, the four mechanisms constitute the complete mechanism system of AI-empowered ICH digital revitalization.

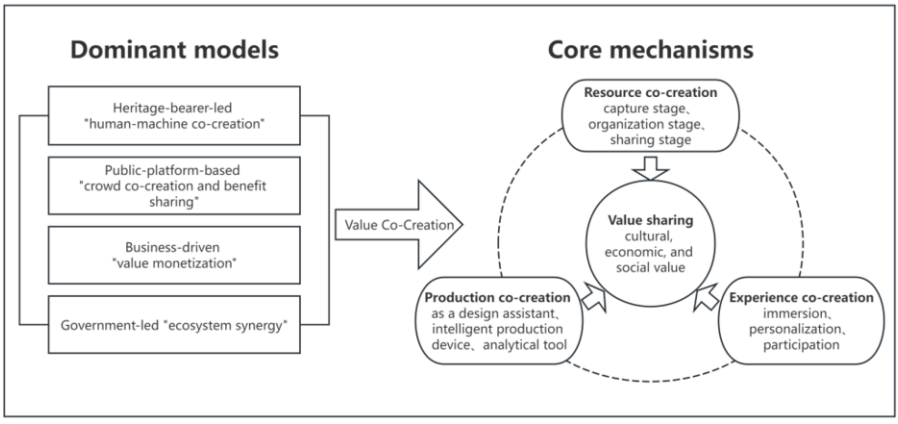


Figure 1. Analytical framework of AI-empowered value co-creation for ICH digital revitalization, including four dominant models and four interactive core mechanisms.

4.2.1. Resource co-creation

Resource co-creation uses AI to capture, structure, and openly aggregate ICH resources, converting tacit knowledge into explicit knowledge and scattered resources into public resources. In the capture stage, high-precision sensing and multimodal recording preserve bearers' movements and sensory judgments, turning face-to-face knowledge into recordable digital signals. In the organization stage, machine learning automatically classifies, labels, and links large volumes of images, texts, audio, and video, building semantically connected ICH knowledge graphs. In the sharing stage, open mechanisms let bearers, researchers, designers, and the public retrieve and reuse resources on demand. The Quanzhou string puppetry team, for example, systematically captured knowledge of the puppet schools and organized the resources into an ICH library. The team then built the Ouyi Yuncheng platform, whose AI-assisted learning and video-based intelligent

evaluation systems let the public learn about puppetry, use the data, and practice. The Longjing Digital Archive and the Shibing digital gene bank illustrate the same chain from digital capture to data organization and resource sharing.

4.2.2. Production co-creation

Production co-creation occurs when AI enters the production and reproduction stage, raising efficiency and converting outputs through human-machine collaboration. AI plays three roles. As a design assistant, it generates culturally consistent design proposals from the resource database built in the previous stage, which bearers and designers can screen and refine. As an intelligent production device, it makes small-batch, personalized customization feasible. As an analytical tool, it reads consumer preferences and market trends to inform ICH product development. The Shu embroidery case validates this chain: AI-generated patterns were converted into personalized products through flexible production. The Dong brocade case goes further. A Hunan University team built a domain-specific model combining small samples with rule constraints and developed intelligent brocade design software. Using 3D scanning, the team turned Dong brocade patterns, structures, and colors into reproducible digital models. The system shortened workflows, lowered the threshold of creation, and directly generated employment for more than 3,000 weavers. AI empowerment thus reshapes not only efficiency but also product forms and market space.

4.2.3. Experience co-creation

Experience co-creation occurs when AI enters the consumption and experience stage, drawing users into ICH transmission, safeguarding, and value creation. AI works here at three levels. First, immersion: VR, AR, and digital humans create ICH scenes beyond time and space, letting audiences enter ancient kiln sites, talk with historical figures, and witness rituals. Second, personalization: AI tailors content to audience preferences and interaction behavior. The agents in the Baidu Baike AI ICH Pavilion, for instance, support knowledge Q&A, multi-turn interaction, and generative experiences, replacing uniform displays with experiences tailored to each user. Third, participation: generative AI removes nearly all technical barriers, so the public can design paper-cutting patterns, generate Kunqu opera accompaniment with one click, and learn string puppetry with the help of digital humans. The Houma shadow puppetry case shows that experience co-creation turns audiences from bystanders into participants. Through deep involvement, audiences internalize the cultural meaning of ICH and identify with it.

4.2.4. Value sharing

Value sharing means that cultural, economic, and social value emerge together from coordinated interaction among actors rather than being captured by any one of them. Cultural value lies in the creative transformation of ICH symbols. Bearers gain broader cultural recognition through co-creation with AI, firms' technological practices take on cultural meaning, and the public deepens its cultural understanding and identity. Government, in turn, achieves tangible governance outcomes through effective dissemination. Economic value lies in the market circulation of ICH products and services. Bearers expand sales channels and build brands, firms earn reasonable returns, and the public obtains quality products and services. Government benefits from industrial spillover that grow local cultural industries. Social value lies in the breadth of

dissemination, the depth of identification, and the degree of participation. Bearers gain wider attention and support, firms strengthen their social responsibility image, and the public builds cultural confidence. Government, in turn, improves social benefits in public cultural services. In the Shibing Miao embroidery case, cultural, economic, and social value emerged together within one institutional ecosystem. Value sharing is not a zero-sum distribution; it is what makes ICH digital revitalization sustainable.

5. Discussion

Three contributions stand out. First, the study extends value co-creation theory. Prior work mostly starts from firm-customer interaction. This study moves the analysis to AI-empowered ICH digital revitalization and identifies a multi-actor service ecosystem of bearers, public platforms, firms, government, and the public. It translates the asymmetry identified by Song et al. (2022) [13] into four dominant models. Second, the study refines the role of technology. Where Jian et al. (2016) [18] treat technology as an operant resource, this study shows that AI has become a digital partner that learns, generates, and collaborates. Through its autonomy, learning capacity, and generativity, AI is fundamentally changing human-machine relations. This is consistent with international evidence that AI enables and enhances multi-actor value co-creation [19], while also warning of possible value co-destruction [20]. In practice, excessive commercialization may erode the cultural meaning of ICH, and algorithmic generation may distort its cultural connotations and traditional aesthetics. Because this study centers on positive value co-creation, these destructive pathways are not developed in depth. Third, the study proposes a two-layer framework of dominant models and core mechanisms that integrate the resource, production, experience, and value stages. This fills the gap on process mechanisms and provides a conceptual tool for future empirical work.

Governments should strengthen institutional frameworks and digital infrastructure, coordinate multi-actor participation, and prevent data rights and algorithmic ethics risks in line with international AI ethics governance. Heritage bearers should uphold cultural subjectivity and decision-making authority, treating AI as an assistant for knowledge while keeping technology from alienating cultural authenticity. Firms and platforms should balance market and cultural logic and resist excessive commercialization. Public platforms and the public should keep data and creation tools open and raise digital literacy so that crowd co-creation and benefit sharing can take root.

6. Conclusion and future directions

This study examines how AI empowers ICH digital revitalization. Drawing on value co-creation theory, it conducts a conceptual multi-case analysis of online public cases and builds and tests a two-layer framework of dominant models and core mechanisms. The findings show that, when AI is introduced, ICH value co-creation takes four forms: heritage-bearer-led human-machine co-creation, public-platform-based crowd co-creation and benefit sharing, business-driven value monetization, and government-led ecosystem synergy. Resource, production, and experience co-creation and value sharing operate together. The study offers a conceptual tool for understanding value creation in ICH in the AI era and practical guidance for actors engaged in ICH digital revitalization.

Three limitations qualify these conclusions. First, reliance on publicly available sources may limit insight into internal mechanisms. Second, the cases cannot capture the full variation of cultural ecosystems. Third, findings are time-bound and will evolve with AI and multimodal advances. Future research should collect first-hand fieldwork data and validate the framework with quantitative and mixed methods.

References

[1] Xi Jinping stressed in the 17th collective study of the Political Bureau of the CPC Central Committee: An choring the strategic goal of building a cultural power and constantly developing socialist culture with C hinese characteristics in the new era. Renmin Ribao. 2024 Oct 29;1.

[2] UNESCO. Convention for the Safeguarding of the Intangible Cultural Heritage; Charter on the Preservati on of Digital Heritage. Paris: UNESCO; 2003.

[3] Song J, Wang M. An analysis of the current situation and problems of the digital protection of Chinese in tangible cultural heritage. Cult Herit J. 2015;(6):1-9.

[4] Wen W, Zhao M. Constructing the digital scenes of intangible cultural heritage in China. Theory Mon. 20 22 Oct;(10):89-99. doi:10.14180/j.cnki.1004-0544.2022.10.011.

[5] Vial G. Understanding digital transformation: a review and a research agenda. J Strateg Inf Syst. 2019 Ju n;28(2):118-44. doi:10.1016/j.jsis.2019.01.003.

[6] Hou Y, Kenderdine S, Picca D, Egloff M, Adamou A. Digitizing intangible cultural heritage embodied: st ate of the art. J Comput Cult Herit. 2022 Sep;15(3):1-20. doi:10.1145/3494837.

[7] Huang Y. The protection and utilization of the intangible cultural heritage in the context of digitization. C ult Herit J. 2015;(1):1-10.

[8] Tan G, He Q. Research status, realistic dilemmas, and development paths of digital communication of int angible cultural heritage in China. Theory Mon. 2021 Sep;(9):87-94. doi:10.14180/j.cnki.1004-0544.20 21.09.010.

[9] Prahalad CK, Ramaswamy V. Co-creation experiences: the next practice in value creation. J Interact Mar k. 2004;18(3):5-14. doi:10.1002/dir.20015.

[10] Vargo SL, Lusch RF. Evolving to a new dominant logic for marketing. J Mark. 2004 Jan;68(1):1-17. do i:10.1509/jmkg.68.1.1.24036.

[11] Vargo SL, Lusch RF. From repeat patronage to value co-creation in service ecosystems: a transcending conceptualization of relationship. J Bus Mark Manag. 2010 Oct;4(4):169-79. doi:10.1007/s12087-010-0 046-0.

[12] Vargo SL, Lusch RF. Institutions and axioms: an extension and update of service-dominant logic. J Aca d Mark Sci. 2016 Jan;44(1):5-23. doi:10.1007/s11747-015-0456-3.

[13] Song X, Liang X, Zhang X, et al. Intangible cultural heritage into scenic spots: multi-case study of the l ogic and mechanism of multi-subject value co-creation. Tourism Tribune. 2022 Nov;37(11):85-100. do i:10.19765/j.cnki.1002-5006.2022.00.015.

[14] Ma X, Chen X. Value co-creation mechanism and activation pathways for intangible cultural heritage to urism: a case study of Jiangyong Nushu. Geogr Geo-Inf Sci. 2025;41(6):138-47.

[15] Dwivedi YK, Hughes L, Ismagilova E, Aarts G, Coombs C, Crick T, et al. Artificial intelligence (AI): multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. Int J Inf Manag. 2021 Apr;57:101994. doi:10.1016/j.ijinfomgt.2019.08.002.

[16] Huang MH, Rust RT. Artificial intelligence in service. J Serv Res. 2018 May;21(2):155-72. doi:10.1177 /1094670517752459.

[17] Chandra B, Rahman Z. Artificial intelligence and value co-creation: a review, conceptual framework an d directions for future research. J Serv Theory Pract. 2024 Jan;34(1):7-32. doi:10.1108/JSTP-03-2023-0 097.

[18] Jian Z, Linghu K, Li L. The evolution and prospects of value co-creation research: a perspective from c ustomer experience to service ecosystems. Foreign Econ Manag. 2016 Sep;38(9):3-20. doi:10.16538/j.c nki.fem.2016.09.001.

[19] Leone D, Schiavone F, Appio FP, Chiao B. How does artificial intelligence enable and enhance value c o-creation in industrial markets? An exploratory case study in the healthcare ecosystem. J Bus Res. 202 1 Jun;129:849-59. doi:10.1016/j.jbusres.2020.11.008.

[20] Bui HT, Filimonau V, Sezerel H. Exploring value co-creation and co-destruction between consumers & generative artificial intelligence (GAI) in travel. Tour Manag Perspect. 2025 Jul;58:101392. doi:10.101 6/j.tmp.2025.101392.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261088

The Happiness Backfire: Engagement-Oriented Brand Content Predicts Fewer Reposts and Shapes Comment Responses

Bowei Sun[a], Yanshu Wang[a] and Tong Li[b,1]
[a] *School of Humanities and Social Science, The Chinese University of Hong Kong, Shenzhen, China*
[b] *School of Business, Guangdong University of Foreign Studies, Guangzhou, China*
ORCiD ID: Bowei Sun https://orcid.org/0009-0005-0903-9598

Abstract. Brand-owned social media increasingly carries messages about enjoyment, participation, connection, purpose, and achievement, yet the same message may not produce the same response. This study examines how engagement-oriented brand content relates to comments and reposts across two brand contexts. Using PERMA as a semantic coding standard, we analyze 1,191 official Weibo posts from McDonald's and Burberry in 2025. A Python-assisted LLM pipeline provides pre-annotation, followed by two-researcher reviews. OLS and PPML show a robust negative association between engagement-oriented content and reposting, while the comment association is less stable. Brand-specific results show greater heterogeneity for reposting. We define this as a narrow happiness backfire: content built around active involvement can be less portable across users' networks even when it invites direct response. The findings show that conversation and public diffusion are distinct communication goals, so content designed to invite participation should not be assumed to travel widely.

Keywords. brand well-being; engagement-oriented content; comments; reposts; PERMA; large language model; Weibo

1. Introduction

Brand-owned social media is a routine setting for brand-consumer interaction. Posts can invite consumers to enjoy experiences, join activities, connect with others, pursue purpose, or celebrate achievement. Yet firms often evaluate these messages through aggregate engagement metrics. This can be misleading because comments and reposts represent different behaviors. Commenting is a local response within the focal post, whereas reposting moves the message into the user's own network and makes the user part of its distribution [1,2]. The question is therefore not only whether well-being-oriented content generates engagement, but what kind of engagement it generates.

Communication channels shape what people transmit [3], public contribution carries image-related utility [4], and viral diffusion depends on whether content retains value outside its original context [5]. Reposting can also make brand association more identity-

[1] Corresponding Author: Tong Li, School of Business, Guangdong University of Foreign Studies, Guangzhou, China; E-mail: Tongi@gdufs.edu.cn.

relevant [6]. Engagement-oriented content is therefore ambiguous: invitations to participate, try, discuss, or experience may make direct response easier, while context-specific value may make public sharing less attractive. Brand social-media research often treats positive content broadly. We instead use PERMA [7] to classify message content and connect this framework to research on brand purpose and eudaimonic well-being [8-10].

We examine 1,191 official Weibo posts from McDonald's and Burberry in 2025. PERMA engagement (E) is the focal construct. A Python-assisted LLM pipeline provides pre-annotation, followed by two-researcher reviews. We estimate comments and reposts separately with OLS and PPML and compare brand-specific E slopes. Engagement-oriented content consistently predicts fewer reposts, while its association with comments is positive in OLS but not robust in PPML. Repost relationships also differ more clearly across the two accounts. Because each category is represented by one brand, these differences are sample-specific. The findings identify a channel-specific limit to well-being-oriented brand communication and show why brands should distinguish content for conversation from content intended for wider diffusion.

2. Theory and Hypotheses

2.1. Literature Review and Research Gap

Social-media engagement is behaviorally heterogeneous. Brand content predicts different forms of response, and linguistic style can relate differently to likes, comments, and shares [11]. Word-of-mouth research also shows that communication channels shape what people transmit. Virality depends on message properties and portable value, not attention alone [12,13], and positive emotion can backfire in some settings [14]. Together, these studies suggest that one content feature needs not move comments and reposts in the same direction.

A parallel literature links brand purpose and consumer well-being to marketing communication. Brand purpose can support eudaimonic well-being, brands communicate higher-purpose themes on social media, and eudaimonic narratives can change product valuation. Automated text analysis is also central to large-scale content coding [15]. LLMs can improve annotation efficiency [16], but recent work supports human validation [17]. The remaining gap is to combine a multidimensional well-being taxonomy with channel-specific outcomes and an auditable human-LLM coding process.

2.2. Engagement-Oriented Well-Being and Channel Divergence

PERMA was developed to describe human well-being. Here it is used only as a semantic taxonomy of brand messages. P denotes positive emotion; E denotes active involvement, absorption, challenge, participation, or experience; R denotes relationships; M denotes meaning; and A denotes accomplishment. N indicates that no qualifying P/E/R/M/A expression is present. This design classifies what a post expresses without claiming that a brand or consumer possesses a PERMA state.

Engagement-oriented posts often present something to do, try, discuss, or experience. That structure can supply a ready object for commenting. Reposting imposes another requirement: the content must also be worth carrying into the user's own network. Written transmission is shaped by channel considerations, image-related motives, and

identity signaling. Public endorsement cost is therefore a plausible explanation for weaker diffusion, but it is not directly measured in this study.

H1a: Compared with brand posts without engagement-oriented expression, posts containing engagement-oriented well-being expression are associated with more consumer comments.

H1b: Compared with brand posts without engagement-oriented expression, posts containing engagement-oriented well-being expression are associated with fewer consumer reposts.

2.3. Brand Type as a Boundary Condition

Brand context may change how an engagement-oriented message is interpreted. Luxury consumption is especially identity-relevant [18], while brand meaning can become part of consumer self-construal [19]. McDonald's and Burberry also differ in audience, usage occasions, campaign style, and symbolic positioning. These differences justify testing whether the E-response slope differs across brands, but they do not identify a general causal effect of brand category.

H2a: In this two-brand sample, brand type moderates the association between engagement-oriented well-being expression and consumer comments.

H2b: In this two-brand sample, brand type moderates the association between engagement-oriented well-being expression and consumer reposts.

3. Research Design

3.1. Sample and Coding Reliability

PERMA coding followed a Python-assisted human-LLM workflow. We first fixed a multilabel handbook for P/E/R/M/A/N, with positive and negative examples, boundary cases, promotional-rhetoric rules, and the residual N rule. A Python script called DeepSeek-V4-Flash through the API with reasoning mode disabled. The fixed prompt supplied label definitions and human-coded few-shot examples and instructed the model to judge post text rather than later engagement outcomes. Posts were submitted in batches of 20. Each response returned six binary labels, a short rationale, and a confidence score in structured JSON. This kept semantic classification separate from later statistical analysis and applied the same decision rules across batches.

The Python pipeline saved raw outputs and checkpoints and checked post IDs, duplicates, missing outputs, legal binary values, and residual-N logic before human review. These checks detected malformed or inconsistent records but did not overwrite labels. We used 200 human-coded reference posts to stabilize the prompt and supply few-shot examples. Comparing model output with final human-reviewed codes produced exact six-label agreement of .732 and micro-F1 of .749, with substantial variation across labels. The expanded Burberry corpus used the same frozen definitions and output schema. Because it was added after the original diagnostic exercise, these agreement values describe the assisted-coding procedure rather than a fresh reliability estimate for every post.

Two researchers reviewed final labels against each complete post and the shared handbook. They could correct any label, with special attention to model-human disagreements, low-confidence cases, and posts combining several PERMA themes.

When judgments differed, they returned to the operational definition and full post context; genuinely ambiguous expressions were coded conservatively as absent. Only reviewed labels entered the analysis. The final corpus contains 1,191 official-account Weibo posts from 2025: 758 from McDonald's and 433 from Burberry, all with nonmissing comment and repost counts. Table 1 reports the sample and final PERMA distribution. The LLM therefore served as pre-annotation rather than autonomous ground truth.

Table 1. Sample and PERMA-label distribution

Brand	Source N	Analytic N	P	E	R	M	A	N
McDonald's	758	758	81 (10.7%)	284 (37.5%)	26 (3.4%)	4 (0.5%)	7 (0.9%)	401 (52.9%)
Burberry	433	433	18 (4.2%)	27 (6.2%)	19 (4.4%)	18 (4.2%)	6 (1.4%)	356 (82.2%)

Note. PERMA labels are nonexclusive; N denotes no qualifying P/E/R/M/A expression.

3.2. Variables and Models

The outcomes are comment and repost counts. The focal predictor is E. Baseline OLS models use ln(1+y) to retain zero outcomes while limiting scale compression from extreme counts. PPML models the raw counts and provides a complementary specification for nonnegative outcomes [20]. All models reported in Table 3 include month fixed effects and robust standard errors. Brand heterogeneity is tested with E × Burberry and brand-specific E slopes. The Wald test has one degree of freedom and therefore tests only whether E × Burberry equals zero; it is not a joint test of several interactions.

3.3. Descriptive Statistics

Social-media outcomes are strongly right-skewed, so Table 2 reports zero rates, means and standard deviations, medians and interquartile ranges, upper percentiles, and full ranges by brand. E is far more common at McDonald's (284/758, 37.5%) than at Burberry (27/433, 6.2%); 284 of the 311 E posts (91.3%) come from McDonald's. Meaning is more frequent at Burberry (18/433, 4.2%) than at McDonald's (4/758, 0.5%). This imbalance matters when brand-specific E slopes are compared.

Table 2. By-brand outcome distributions

Brand	Outcome	N	Zero %	Mean (SD)	Median [P25–P75]	P95; P99	Range
McDonald's	Comment	758	5.1	304.46 (512.96)	141.00 [50.25–395.50]	1032.85; 2084.38	0–8911
McDonald's	Reposts	758	8.0	317.00 (1341.70)	45.00 [11.00–136.75]	1185.45; 4982.10	0–22921
Burberry	Comment	433	0.2	522.40 (1009.81)	75.00 [18.00–556.00]	2360.40; 4622.28	0–10243
Burberry	Reposts	433	0.7	1141.01 (4445.75)	69.00 [17.00–563.00]	3599.20; 18627.44	0–52826

4. Empirical Results

4.1. Main and Brand-Specific Effects

Table 3 gives a mixed result for H1a. In the pooled sample, E is positively associated with comments in OLS (b=.363, p=.002), but the PPML estimate is essentially zero (b=.009, p=.924). H1b is much more stable: E predicts fewer reposts in both OLS (b=-1.077, p<.001) and PPML (b=-1.894, p<.001). The channel-divergence claim is therefore strongest for reposting, while the comment result depends on estimator and context.

Brand-specific estimates clarify that instability. For McDonald's comments, E is positive in OLS (b=.383, p=.003) and PPML (b=.350, p=.002). Burberry's comment slopes are also positive but imprecise (OLS b=.589, p=.121; PPML b=.167, p=.519); the E × Burberry interaction is nonsignificant in both models, so H2a is not supported as a formal slope difference. Reposts show the opposite moderation result. McDonald's E slope is strongly negative, whereas Burberry's slope is weak and model-dependent. The repost interaction is significant in OLS (b=1.649, p<.001) and PPML (b=1.074, p=.010), supporting H2b within this two-brand sample. The Wald tests are one-degree-of-freedom tests of that same interaction coefficient.

Table 3. Engagement-oriented-content estimates with month fixed effects

Test	Comments OLS	Comments PPML	Reposts OLS	Reposts PPML
Pooled E	0.363 (0.120), .002	0.009 (0.094), .924	−1.077 (0.119), <.001	−1.894 (0.217), <.001
McDonald's E	0.383 (0.129), .003	0.350 (0.115), .002	−1.028 (0.130), <.001	−1.665 (0.267), <.001
Burberry E	0.589 (0.380), .121	0.167 (0.259), .519	0.621 (0.363), .087	−0.591 (0.327), .071
E × Burberry	0.205 (0.401), .609	−0.183 (0.283), .517	1.649 (0.386), <.001	1.074 (0.419), .010
Wald $\chi^2(1)$, p	0.26, .609	0.42, .517	18.26, <.001	6.56, .010

Note. Entries are b (robust SE), p unless stated otherwise; N = 1,191. "Pooled E" comes from the no-interaction model. Each Wald test evaluates only E × Burberry = 0 (equal brand-specific E slopes), not joint significance of multiple interactions.

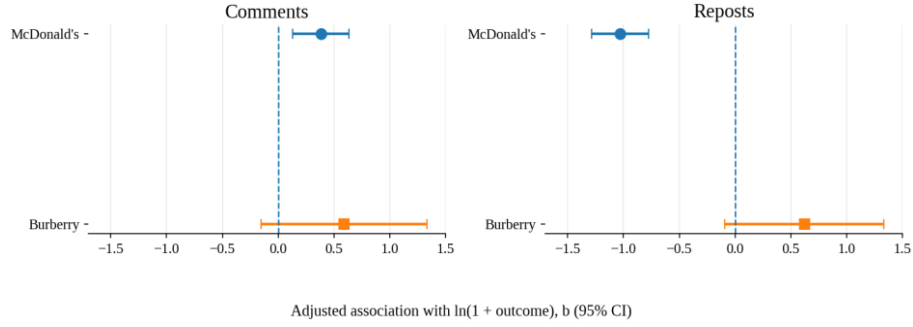


Figure 1. Brand-specific E coefficients in OLS models with month fixed effects. Error bars are 95% confidence intervals; the dashed line denotes zero.

4.2. Robustness Checks

Table 4 shows that the main qualitative pattern is not driven by a few viral posts. For comments, McDonald's E slope remains positive after trimming, winsorization, and week-clustered inference. Burberry's comment slope is positive but less precise, and the interaction remains nonsignificant. For reposts, McDonald's negative E slope remains

highly significant in every check. Burberry's OLS slope is positive or imprecise, and the brand interaction remains significant. Thus the pooled E–repost association is robust, but its magnitude is not uniform across the two brands.

Table 4. Selected robustness checks

Outcome	Check	McDonald's E b, p	Burberry E b, p	Interaction p
Comments	Trim top 1%	0.402, .002	0.696, .070	.467
Comments	Winsorize P99	0.386, .003	0.599, .115	.596
Comments	Month FE, week cluster	0.383, .013	0.589, .272	.723
Reposts	Trim top 1%	−0.995, <.001	0.731, .043	<.001
Reposts	Winsorize P99	−1.026, <.001	0.631, .082	<.001
Reposts	Month FE, week cluster	−1.028, <.001	0.621, .159	<.001

5. Discussion and Conclusion

5.1. Conclusion

The results support a narrow form of happiness backfire. Engagement-oriented content predicts fewer reposts in pooled OLS and PPML, and the result survives trimming and winsorization. Comments are less stable: OLS is positive, PPML is not, and the Burberry estimate is less precise. The formal comment interaction is nonsignificant. Thus, lower diffusion is the robust result, whereas the comment response is context-sensitive.

This asymmetry fits the different demands of the two channels. Commenting responds to the focal post; reposting carries the message into the user's own network, where portability, audience fit, and identity presentation matter more. Image-related utility and identity signaling also shape public contribution. Public endorsement cost is therefore plausible but not directly measured.

5.2. Contributions, Limitations, and Future Research

The study makes three contributions. First, it applies PERMA as a message-level taxonomy and isolates engagement-oriented expression rather than treating positive content as one category. Second, it separates direct response from public diffusion, showing why aggregate engagement measures can mislead. Third, it combines a documented Python-LLM pre-annotation pipeline with two-researcher review and robustness checks, making content measurement more scalable while keeping final labels human-reviewed and auditable.

Several limitations remain. The study is observational, single-platform, and limited to two brands. E is unevenly distributed (27 Burberry vs. 284 McDonald's posts), reducing common support and Burberry precision. The expanded corpus used the same codebook, but a fresh independent double-coding reliability estimate was not reconstructed for all 1,191 posts. Public endorsement cost is inferred rather than measured. Future work should add brands and platforms, retain independent coder files and held-out validation, compare LLMs and prompts, and experimentally separate comment invitations from shareable content.

References

[1] Lee D, Hosanagar K, Nair HS. Advertising content and consumer engagement on social media: evidence from Facebook. Manage Sci. 2018;64(11):5105-5131.
[2] Dhaoui C, Webster CM. Brand and consumer engagement behaviors on Facebook brand pages: Let's have a (positive) conversation. Int J Res Mark. 2021;38(1):155-175.
[3] Berger J, Iyengar R. Communication channels and word of mouth: how the medium shapes the message. J Consum Res. 2013;40(3):567-579.
[4] Toubia O, Stephen AT. Intrinsic vs. image-related utility in social media: why do people contribute content to Twitter? Mark Sci. 2013;32(3):368-392.
[5] Berger J, Milkman KL. What makes online content viral? J Mark Res. 2012;49(2):192-205.
[6] Berger J, Heath C. Where consumers diverge from others: identity signaling and product domains. J Consum Res. 2007;34(2):121-134.
[7] Seligman MEP. Flourish: A Visionary New Understanding of Happiness and Well-Being. New York: Free Press; 2011.
[8] Williams P, Escalas JE, Morningstar A. Conceptualizing brand purpose and considering its implications for consumer eudaimonic well-being. J Consum Psychol. 2022;32(4):699-723.
[9] Rudeloff C, Michalski P. How corporate brands communicate their higher purpose on social media: evidence from top global brands on Twitter. Corp Reput Rev. 2024;27:202-215.
[10] Hamby A, Tezer A, Escalas JE. Significant objects: how eudaimonic narratives enhance the value of featured products. J Advert. 2023;52(3):406-422.
[11] Deng Q, Hine MJ, Ji S, Wang Y. Understanding consumer engagement with brand posts on social media: the effects of post linguistic styles. Electron Commer Res Appl. 2021;48:101068.
[12] Tellis GJ, MacInnis DJ, Tirunillai S, Zhang Y. What drives virality of online digital content? J Mark. 2019;83(4):1-20.
[13] Akpinar E, Berger J. Valuable virality. J Mark Res. 2017;54(2):318-330.
[14] Rocklage MD, Fazio RH. The enhancing versus backfiring effects of positive emotion in consumer reviews. J Mark Res. 2020;57(2):332-352.
[15] Humphreys A, Wang RJH. Automated text analysis for consumer research. J Consum Res. 2018;44(6):1274-1306.
[16] Gilardi F, Alizadeh M, Kubli M. ChatGPT outperforms crowd workers for text-annotation tasks. Proc Natl Acad Sci USA. 2023;120(30):e2305016120.
[17] Pangakis N, Wolken S. Keeping humans in the loop: human-centered automated annotation with generative AI. Proc Int AAAI Conf Web Soc Media. 2025;19(1):1471-1492.
[18] Han YJ, Nunes JC, Dreze X. Signaling status with luxury goods: the role of brand prominence. J Mark. 2010;74(4):15-30.
[19] Escalas JE, Bettman JR. Self-construal, reference groups, and brand meaning. J Consum Res. 2005;32(3):378-389.
[20] Santos Silva JMC, Tenreyro S. The log of gravity. Rev Econ Stat. 2006;88(4):641-658.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)

© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261089

How Brand Happiness Spreads in User Co-creation — Evidence from Large-Scale Social Media Data

Guo TANG[a], Yongtian YU[a] and Tong LI[b,1]
[a] *School of Humanities and Social Science, The Chinese University of Hong Kong, Shenzhen, China*
[b] *School of Business, Guangdong University of Foreign Studies, Guangzhou, China*
ORCiD ID: Guo Tang https://orcid.org/0009-0007-5245-2753

Abstract. User co-creation has transformed digital marketing, yet how happiness-related content relates to brand engagement in co-creation contexts remains understudied. Integrating user co-creation, brand happiness, and content diffusion theories, this study empirically examines the happiness–engagement association using 51,642 posts from 232 brand-related topics on Weibo (January 2025–January 2026). Three findings emerge. (1) Happiness perception in co-created content is positively associated with brand engagement ($\beta = 0.206$, $p = 0.014$, IRR = 1.229); a one-SD increase in happiness corresponds to approximately 12%–13% higher engagement. (2) Creative expression co-creation shows a positive but non-significant moderating trend in the primary Poisson model ($p = 0.571$), though supplementary OLS analysis suggests a potential pattern ($p = 0.025$). (3) Exploratory subgroup analysis indicates the happiness–engagement association is significant for FMCG brands ($\beta = 0.231$, $p < 0.01$) but not for gaming brands ($\beta = 0.089$, n.s.), though the pooled interaction test is non-significant ($p = 0.404$). This study is the first to systematically examine brand happiness engagement from a user co-creation perspective.

Keywords. User co-creation; brand happiness; social media engagement; LIWC dictionary; Poisson regression

1. Introduction

Value co-creation — the collaborative process through which firms and consumers jointly generate brand meaning, products, and experiences — has become a central strategic imperative[1]. Traditional branding followed a firm-controlled logic, broadcasting professionally crafted messages to passive audiences. Participatory branding has disrupted this paradigm: brands are no longer firm monologues but ongoing conversations co-authored with audiences[2]. Social media enable consumers to generate vast quantities of user-created content (UGC) — reviews, brand challenges, consumption stories — that shapes brand narratives in real time[3,4].

[1] Corresponding Author: Tong Li, School of Business, Guangdong University of Foreign Studies, Guangzhou, China; E-mail: Tongi@gdufs.edu.cn.

Within this landscape, happiness has emerged as among the most resonant emotional themes. Users routinely express joy, warmth, and satisfaction in brand-related posts — experiences central to subjective well-being[5,6] — and such happiness-laden content enjoys a privileged position in social platforms' attention economy[7,8]. Brand happiness, the positive affective state elicited by brand experiences[9], is both personally meaningful and socially shareable. As co-creators, users imbue brands with emotional significance through topic challenges, campaigns, and creative reinterpretations[3,4], generating shared affective experiences that resonate across brand communities[1,2].

Yet whether happiness in user co-created content is associated with broader brand engagement — and under what conditions — remains empirically untested. Existing research examines co-creation, brand happiness, and emotional diffusion largely in isolation. First, co-creation studies focus on functional and economic value, neglecting affective dimensions[2,10]. Second, brand happiness research treats happiness as an internal consumption outcome[9], not as a content feature that propagates socially. Third, emotion diffusion research shows high-arousal content spreads more readily[8,11], but rarely examines co-creation-specific boundary conditions.

This study asks: How is happiness perception in user co-created content associated with brand engagement, and how do co-creation value type and brand type relate to this association? Using 51,642 posts from 232 brand-related topics on Weibo over 13 months (January 2025–January 2026), we examine whether the happiness–engagement association varies by (a) creative expression degree and (b) brand category (gaming, FMCG, luxury). The study contributes by: (1) reconceptualizing brand happiness from internal outcome to socially shared content feature; (2) integrating co-creation theory with brand happiness and diffusion research; and (3) providing exploratory evidence on boundary conditions — value type and brand type — that may shape the happiness–engagement relationship.

2. Literature Review and Hypothesis Development

2.1. User Co-creation, Brand Happiness, and Brand Engagement

Value co-creation theory holds that consumers co-construct brand identity through UGC[1,2,3,4]. While this literature has documented functional dimensions including product innovation[10] and community-based value creation[3], it has paid less attention to affective value in co-creation. Happiness is rarely examined as a co-created content feature with downstream brand consequences.

Brand happiness refers to the positive affective state elicited by brand experiences, characterized by joy, warmth, and contentment[9]. It is distinct from brand love, which emphasizes deep attachment and passion[12,13], and from brand attachment, which reflects the cognitive-emotional consumer–brand bond[14]. Brand happiness captures a more accessible form of positive affect readily expressed in everyday user-generated content and shared through social interaction, making it particularly suited for studying affective dynamics in co-creation.

Social media engagement — the aggregate of likes, comments, and reposts — captures both diffusion breadth and interaction depth[15]. High-arousal positive content is more likely to be shared[8], supported by broaden-and-build theory, which posits that positive emotions expand cognitive repertoires and motivate prosocial

behaviors including sharing[11]. Happiness should therefore be associated with higher engagement in co-creation contexts.

H1: Happiness perception in user co-created content is positively associated with brand engagement.

2.2. Boundary Role of Co-creation Value Type

Co-created content varies in value type — functional, affective, and creative[3,4]. Creative expression, involving novel ideas and imaginative brand reinterpretations[3,16], may strengthen happiness-driven engagement through multiple pathways.

First, co-creation enhances happiness through psychological ownership and intrinsic motivation: actively shaping brand narratives elevates positive affect[4]. Creative expression provides opportunities for self-actualization, amplifying hedonic experience[3], while peer recognition reinforces positive emotions through reciprocal affirmation in brand communities[16].

Second, once happiness is present in creative content, novelty-enhanced arousal may amplify its diffusion potential[8,17]. The combination of happiness warmth and novelty surprise may confer higher social currency[7], making such content more shareable. However, the small proportion of highly creative content may attenuate statistical detectability of this moderation effect.

H2: Creative expression co-creation is associated with a stronger positive relationship between happiness and engagement.

2.3. Boundary Role of Brand Type

Brand type may shape the happiness–engagement relationship. FMCG brands integrate with daily life, creating close psychological distance and broad reach[18,19], while gaming brands center on entertainment. Three mechanisms suggest stronger happiness–engagement associations for FMCG: closer emotional contagion[19,20], wider diffusion networks[18], and greater perceived authenticity[21].

Emotional contagion theory supports this boundary condition. Positive emotions spread more rapidly when psychological distance is low[20]. FMCG brands, embedded in everyday routines, may facilitate closer emotional resonance, amplifying the happiness–engagement association.

H3: FMCG brands show a stronger positive relationship between happiness and engagement than experience-oriented brands (e.g., gaming brands).

3. Research Design

3.1. Data Source and Sample

This study uses China's Weibo platform. Through keyword search and topic filtering, we collected posts covering gaming, FMCG, and luxury brands over 13 months (January 2025–January 2026).

The raw dataset contains 121,778 brand-related posts covering 232 topic keywords. After data cleaning — removing duplicates, pure reposts, marketing accounts, and

excessively short texts — the final sample consists of 51,642 valid original user co-creation posts.

Sample distribution: gaming brands 32,599 posts (63.1%), FMCG brands 13,140 posts (25.4%), luxury brands 5,903 posts (11.4%).

To ensure data quality, we applied multi-stage preprocessing: removing duplicates and pure reposts to retain original content; filtering marketing accounts by screening posting histories; excluding posts below a minimum text length to eliminate spam; and controlling for visibility confounds by including log followers as a control variable.

3.2. Variable Measurement

Dependent variable (total_engagement): Total interactions (likes + comments + reposts), capturing both communication breadth and engagement depth.

Independent variable (happy_word_ratio): Measured via LIWC dictionary method[22] using a custom 28-word Chinese happiness dictionary (e.g., happy, joy, warm, sweet, love, satisfaction). happy_word_ratio is the proportion of happiness words to total words per post (range: 0–1; M = 0.59, SD = 0.59), treating happiness as a continuous construct.

Moderators. (1) Creativity (creative_ratio): proportion of creativity-related words (novelty, design, creativity, inspiration) to total words (M = 0.28, SD = 0.28). Both happy_word_ratio and creative_ratio are proportions, not percentages. (2) Brand type: gaming (reference), FMCG, luxury, coded as dummy variables.

Controls. Log followers, verification status, log text length, weekend posting, and brand fixed effects. All VIF values < 2.0, indicating no severe multicollinearity.

3.3. Analytical Strategy and Model Specification

We use Poisson quasi-maximum likelihood (Poisson QML) as the primary method, which provides consistent estimates with overdispersed count data[23] and directly yields incidence rate ratios (IRR). Topic-level clustered robust standard errors address non-independence of posts under the same keyword.

The baseline model tests the main effect of happiness on engagement; interaction terms test moderation effects; group regressions and simple slope analysis further explore boundary conditions. Robustness checks include OLS, negative binomial, variable substitution, and outlier treatment. OLS results are reported as supplementary analysis; Poisson QML is treated as the primary model throughout.

4. Empirical Results

4.1. Descriptive Statistics

Table 1. Descriptive Statistics and Correlation Coefficients of Key Variables

Variable	Mean	SD	1	2	3	4	5
1. Happy word ratio	0.59	0.59	1				
2. Engagement (log)	2.45	2.31	0.043***	1			
3. Followers (log)	5.64	3.94	0.005	0.511***	1		
4. Text length (log)	4.66	0.80	−0.143***	0.163***	0.267***	1	
5. Creativity	0.28	0.28	−0.405***	—	—	—	1

Note. *** p<0.001, ** p<0.01, * p<0.05. Creativity and happiness are significantly negatively correlated.

Table 1 reports descriptive statistics and correlations. Happiness correlates positively with engagement ($r = 0.043$, $p < 0.001$), preliminarily supporting H1. Happiness and creativity are negatively correlated ($r = -0.405$, $p < 0.001$), suggesting an affective warmth–creativity trade-off.

4.2. Main Effect and Moderation Effect Tests

Table 2. Main Effect and Moderation Effect Test Results (Poisson QML)

Variable	M1 Controls	M2 Main Effect (H1)	M3 Creativity Mod. (H2)	M4 Brand Mod. (H3)	M5 Full Model
Happy word ratio	—	0.2060*	0.2140*	0.0891	0.1002
	—	(0.0840)	(0.0915)	(0.1545)	(0.1603)
	—	[p=0.014]	[p=0.020]	[p=0.564]	[p=0.533]
Creativity	—	—	0.2530**	—	0.2490**
	—	—	(0.1110)	—	(0.1086)
Happy × Creative	—	—	0.3726	—	0.3981
	—	—	(0.6512)	—	(0.6545)
	—	—	[p=0.571]	—	[p=0.542]
FMCG brand	—	—	—	−0.452**	−0.447**
	—	—	—	(0.180)	(0.178)
Luxury brand	—	—	—	−0.623*	−0.618*
	—	—	—	(0.325)	(0.323)
Happy × FMCG	—	—	—	0.1418	0.1463
	—	—	—	(0.1700)	(0.1712)
	—	—	—	[p=0.404]	[p=0.393]
Happy × Luxury	—	—	—	0.1675	0.1702
	—	—	—	(0.2735)	(0.2741)
	—	—	—	[p=0.540]	[p=0.534]
Control variables	Yes	Yes	Yes	Yes	Yes
Brand fixed effects	Yes	Yes	Yes	No	No
IRR (happiness)	—	1.229	—	—	—
Pseudo R^2	0.394	0.402	0.404	0.398	0.401
Sample size	51,642	51,642	51,642	51,642	51,642

Note: Topic-level clustered robust standard errors in parentheses. 95% CI for H1 IRR: [1.042, 1.449]. *** p<0.001, ** p<0.01, * p<0.05.

As shown in Table 2, H1 (M2). Happiness is positively and significantly associated with engagement ($\beta = 0.206$, SE = 0.084, $p = 0.014$, IRR = 1.229, 95% CI [1.042, 1.449]). A one-SD increase in happiness ($e^{(0.206 \times 0.59)} - 1 \approx 12.9\%$) is associated with a meaningful increase in engagement — a medium-to-large effect in social media content research. H1 is supported.

H2 (M3). The happy × creative interaction is positive but non-significant in Poisson QML ($\beta = 0.373$, 95% CI [−0.903, 1.649], $p = 0.571$). Supplementary OLS yields a significant interaction ($\beta = 0.566$, $p = 0.025$), with the happiness–engagement association roughly three times larger at high vs. low creativity (+20.6% vs. +6.6%). The directional consistency suggests a potential pattern, but the non-significant primary interaction precludes firm conclusions. H2 is unsupported by the primary analysis; the OLS result is exploratory.

H3 (M4). The pooled interaction terms are non-significant (Happy × FMCG: $\beta = 0.142$, 95% CI [−0.191, 0.475], $p = 0.404$; Happy × Luxury: $\beta = 0.168$, 95% CI [−0.369, 0.705], $p = 0.540$). Exploratory subgroup regressions show a significant happiness–engagement association for FMCG brands ($\beta = 0.231$, $p < 0.01$) but not for gaming ($\beta = 0.089$, $p = 0.564$) or luxury brands ($\beta = 0.257$, $p = 0.307$). Without a formal test confirming inter-group coefficient differences, H3 is unsupported.

4.3. Robustness Checks

Five robustness checks confirm H1's stability: OLS with log-transformed engagement, negative binomial regression, raw happiness word count as alternative IV, 1% winsorizing, and brand-level clustering. All yield significant positive effects ($\beta = 0.17$–0.30, all $p < 0.05$), with effect sizes comparable to the primary model.

5. Discussion and Conclusion

5.1. Main Findings

This study examines how happiness perception in co-created content relates to brand engagement using 51,642 posts. Three findings emerge.

First, happiness in user co-created content is positively associated with brand engagement (H1 supported; one-SD ≈ 12%–13% increase), consistent across five robustness checks. This extends research on positive emotion and content virality[8,11] into the brand co-creation context.

Second, creative co-creation does not significantly moderate the happiness–engagement association in the primary model (H2 unsupported). Supplementary OLS suggests a potential pattern, but the non-significant interaction precludes firm conclusions. The attenuation likely reflects the small proportion of creative content (3.3%) and the negative happiness–creativity correlation ($r = −0.405$), which reduce statistical power.

Third, brand type does not yield a significant interaction (H3 unsupported). Exploratory subgroup analysis shows a significant association for FMCG but not gaming brands, though formal inter-group comparisons are non-significant.

5.2. Theoretical Contributions and Practical Implications

Theoretical contributions are threefold. First, this study extends co-creation research from functional toward affective value by showing that co-creation generates brand happiness and is associated with its social sharing. Second, it introduces engagement as a downstream outcome of brand happiness, shifting from treating happiness as internal state to understanding its behavioral implications. Third, it provides exploratory evidence on boundary conditions, offering directions for future research.

Practically, brands may benefit from fostering happiness-oriented co-creation through topic challenges and story campaigns. FMCG brands may find such strategies particularly effective, while gaming brands might benefit more from integrating happiness into gameplay contexts.

5.3. Limitations and Future Research

Limitations. First, LIWC cannot detect irony or metaphorical happiness; deep learning methods could improve validity. Second, observational design limits causal inference. Third, single-platform data (Weibo) raises generalizability concerns. Fourth, bots and algorithmic bias may introduce noise. Fifth, small creative content and luxury samples limit moderation detection power.

Future research should explore mechanisms through experimental or longitudinal designs, distinguish happiness types (exciting vs. calm, hedonic vs. eudaimonic), and extend to short video and live streaming. Formal inter-group coefficient tests with balanced samples would strengthen boundary condition claims.

References

[1] Prahalad CK, Ramaswamy V. The future of competition: co-creating unique value with customers. Boston: Harvard Business School Press; 2004. 240 p.
[2] Hatch MJ, Schultz M. Toward a theory of brand co-creation with implications for brand governance. J Brand Manag. 2010;17(8):590-604.
[3] Schau HJ, Muñiz AM, Arnould EJ. How brand community practices create value. J Mark. 2009;73(5):30-51.
[4] Ranjan K, Read S. Value co-creation: concept and measurement. J Acad Mark Sci. 2016;44(3):290-315.
[5] Ahuvia AC. If money doesn't make us happy, why do we act as if it does? J Econ Psychol. 2008;29(4):491-507.
[6] Sirgy MJ. The psychology of quality of life: hedonic well-being, life satisfaction, and eudaimonia. 2nd ed. Dordrecht: Springer; 2012. 584 p.
[7] Berger J. Contagious: why things catch on. New York: Simon & Schuster; 2013. 256 p.
[8] Berger J, Milkman KL. What makes online content viral? J Mark Res. 2012;49(2):192-205.
[9] Schnebelen S, Bruhn M. An appraisal framework of the determinants and consequences of brand happiness. Psychol Mark. 2018;35(2):101-119.
[10] von Hippel E. Democratizing innovation. Cambridge (MA): MIT Press; 2006. 336 p.
[11] Fredrickson BL. The broaden-and-build theory of positive emotions. Philos Trans R Soc Lond B Biol Sci. 2004;359(1449):1367-1378.
[12] Batra R, Ahuvia A, Bagozzi RP. Brand love. J Mark. 2012;76(2):1-16.
[13] Carroll BA, Ahuvia AC. Some antecedents and outcomes of brand love. Mark Lett. 2006;17(2):79-89.
[14] Thomson M, MacInnis DJ, Park CW. The ties that bind: measuring the strength of consumers' emotional attachments to brands. J Consum Psychol. 2005;15(1):77-91.
[15] Hollebeek LD, Glynn MS, Brodie RJ. Demarketering or redefining the conceptual domain of customer brand engagement. J Serv Manag. 2014;25(2):181-201.
[16] Muniz AM, O'Guinn TC. Brand community. J Consum Res. 2001;27(4):412-432.
[17] Petty RE, Cacioppo JT. Communication and persuasion: central and peripheral routes to attitude change. New York: Springer; 1986. 294 p.
[18] Mogilner C, Aaker JL, Kamvar SD. How happiness affects choice. J Consum Res. 2012;39(2):429-443.
[19] Trope Y, Liberman N. Construal-level theory of psychological distance. Psychol Rev. 2010;117(2):440-463.
[20] Hatfield E, Cacioppo JT, Rapson RL. Emotional contagion. Curr Dir Psychol Sci. 1993;2(3):96-100.
[21] Friestad M, Wright P. The persuasion knowledge model: how people cope with persuasion attempts. J Consum Res. 1994;21(1):1-31.
[22] Pennebaker JW, Boyd RL, Jordan K, Blackburn K. The development and psychometric properties of LIWC2015. Austin: University of Texas at Austin; 2015.
[23] Wooldridge JM. Distribution-free estimation of some nonlinear panel data models. J Econom. 1999;90(1):77-97.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261090

Beyond Hashtag Quantity: How Content Types Trigger Customer Inspiration in Short-Video Marketing Differently

Lanxin Xie [a]; Tong Li [b] and Yongtian Yu [c,1]
[a] *School of Humanity and Social Science, The Chinese University of Hong Kong, Shenzhen, China*
[b] *School of Business, Guangdong University of Foreign Studies, China*
[c] *Peking University Shenzhen Graduate School, China*
ORCiD ID: Lanxin Xie https://orcid.org/0009-0002-0371-2890

Abstract. Hashtags are a key component of information tagging and algorithmic recommendations on short-video platforms. Yet existing research in the field of short-video marketing has largely focused on the number of hashtags, while neglecting the differentiated effects that different types of hashtag content have on consumer psychological responses. To address this gap, this study utilized text mining techniques to analyze 999 short videos and 22,980 consumer comments from five beauty brands on the Douyin platform. The results indicate that the impact of hashtags is content-specific, with "product category hashtags" uniquely and significantly stimulating consumers' action inspiration (e.g., purchase intent). The study confirms that action inspiration and conceptual inspiration are complementary rather than substitutes for one another, and that action inspiration effectively predicts downstream sales conversions. Furthermore, tag density plays a key moderating role: in videos with high tag density, the effect of "lifestyle scenario tags" on inspiration shifts from being ineffective to having a strong positive impact. This study refutes the assumptions of the traditional Theory of Explanatory Level (CLT) and proposes that information accessibility, rather than explanatory level, drives this effect, thereby providing practical guidance for brand managers to prioritize the use of product category tags and to activate the potential of scenario tags through combinatorial strategies.

Keywords. Hashtag content, customer inspiration, short-video marketing, construal level theory

1. Introduction

Short-form video platforms have transformed brand marketing, with hashtags serving as micro-stimuli that guide consumer attention. Yet existing research focuses on hashtag quantity [1], overlooking whether content under different types of hashtags puts various impact on consumers' psychological responses. Customer inspiration, defined as a motivational state facilitating the transition from idea reception to goal pursuit [2]—offers a relevant lens, comprising actionable (behavioral) and ideational (cognitive)

[1] Corresponding Author: Yongtian Yu, Peking University Shenzhen Graduate School, China; Email: yuyongtian@celerintelligence.com

dimensions. Despite its significance, customer inspiration has been studied primarily in traditional advertising and retail [1, 3], with limited attention to micro-stimuli in short-video contexts.

Construal Level Theory (CLT) [4] posits that abstract (high-construal) information requires contextual anchors to be transformed into actionable information, whereas concrete (low-construal) content is self-sufficient. Consequently, topic tags can be broadly categorized into (I) concrete tags—that is, tags with practical meaning that describe a specific, existing object, such as brand names (e.g., #Baiyunshan) and product names (e.g., #Mask)—and (II) abstract tags—terms that do not refer to objects but instead describe life situations (e.g., #Mysummerlife).

Another theory relevant to hashtag content is the spreading activation model [5], which posits that co-occurring concepts activate one another through semantic networks, thereby providing activation pathways that make the meanings of abstract hashtags easier to access. Consequently, we argue that hashtag density can influence the activation of semantic networks and, in turn, affect the degree of insight generated. However, this relationship has not yet been empirically tested. There are currently three research gaps: (1) a lack of a classification system for the content types of topic tags; (2) the differentiated pathways through which different tags lead to inspiration have not yet been examined; and (3) a lack of empirical evidence regarding the path from inspiration to sales in short-form video marketing.

Consequently, the content of the hashtags themselves has caught our attention. Bloggers engaged in commercial marketing often use set combinations, such as "brand name, product name." This is done both to meet the requirements of the product's brand and to ensure their videos are targeted more accurately to their intended audience. However, short-form videos offer a great deal of freedom when it comes to hashtags. Beyond brand and product tags, bloggers can freely choose other tags to label their videos. These tags cover a wide range of topics—from daily life and fashion styles to popular values—and essentially represent bloggers' proactive efforts to categorize their own video content.

The study employed text mining methods to measure inspiration and found that product category tags uniquely enhanced action-oriented inspiration. Contrary to the predictions of Construal Level Theory, the construal level did not predict inspiration, suggesting that information accessibility—rather than the construal level—drives this effect. The contribution of this study lies in shifting the research focus from the number of topic tags to content type and establishing the causal pathway linking tags, inspiration, and sales.

2. Theoretical Background & Hypotheses

Understanding how short-video hashtags might stimulate customer inspiration requires grounding in the broader theoretical architecture of customer inspiration research.

Construal Level Theory (CLT), proposed by Trope and Liberman [4], explains how psychological distance influences the way individuals mentally represent objects, events, and decisions. CLT suggests that greater distance leads to higher-level construals characterized by abstract and decontextualized thinking, whereas proximity promotes concrete and detailed representations. Over time, CLT has been widely applied in areas such as decision-making, consumer behaviour, environmental psychology, and risk perception.

Böttger et al. [6] define customer inspiration as a temporary motivational state that transitions a marketing-induced idea into an intrinsic consumption-related goal. They decompose this state into two causally linked components: an "inspired-by" activation state of evocation and transcendence from novel stimuli, and an "inspired-to" intention state where activation becomes the motivated pursuit of a new idea. The conceptualization of inspiration as a distinct psychological construct has been established in foundational work demonstrating its unique motivational properties [7].

Their experimental work on Pinterest-style recipe recommendations demonstrates that inspirational content interacts with pre-existing "idea shopping" motivation to elicit "inspired-by" states, which subsequently mediate downstream purchase intentions and exploratory behavior. Furthermore, they show that this domain-specific measure outperforms general inspiration scales in predictive validity. Consequently, any content-aggregation mechanism—such as hashtags clustering short-video content around a shared theme—can function as an external trigger for the "inspired-by" state, provided the encountered content is visually appealing and cognitively novel rather than repetitive.

Ki et al. [8] extend customer inspiration theory using social defaults theory, showing that influencer traits (attractiveness, credibility, closeness, and interactivity) trigger an "inspired-by" state. This state leads consumers to adopt influencers' lifestyles and recommendations as "social defaults" when facing preference uncertainty. The resulting "inspired-to" state drives both behavioral outcomes—such as product and brand choice imitation—and emotional consequences like a strong affective bond with the source, with closeness showing stronger effects among male consumers. Further evidence demonstrates that customer inspiration leads to concrete behavioral outcomes, including purchase intentions and exploratory behavior, with social media content serving as a key inspirational source [9].

This framework applies directly to hashtag research because short-video hashtags aggregate multiple thematic exemplars much like a curated feed, amplifying the social-proof function of observed choices even when coming from a diffuse crowd of anonymous creators.

Recent work has begun examining the informational content of social tags—showing that tags carry brand-relevant information [10, 11, 12, 13] and enhance e-WOM diffusion [14]. However, no study has classified hashtags by content type or tested their differential effects on consumer psychological outcomes such as inspiration.

Based on theories above, the first hypothesis is proposed:

H1: Different hashtag content will yield different insights.

H1a: Concrete tags positively influence actionable inspiration.

H1b: Abstract tags positively influence ideational inspiration.

H2: There is a sequential mediating effect between conceptual inspiration, actionable inspiration, and sales growth.

H3: Actionable inspiration has a stronger direct effect on sales than the indirect effect of ideational inspiration.

Although the focus of this study has shifted to the content of the tags themselves, existing research indicates that the number of tags also influences marketing outcomes. In social media contexts, studies examining hashtag quantity report mixed findings: some identify an inverted-U relationship between hashtag count and engagement [15, 16], while others find positive linear effects [17].

Therefore, H1 is supplemented by:

H1c: The impact of the number of hashtags on inspiration follows a pattern of first decreasing and then increasing, i.e. too many or too few hashtags reduce their impact, while a moderate number of hashtags yields the best results.

H1d: Tag diversity has a positive impact on inspiration.

3. Data and Methodology

3. 1 Data

We collected 999 short videos from five beauty brands on Douyin (forever key, Wenboshi, Baiyunshan, Zaowuzhe, JCprogram), spanning July 2023 to September 2024. The data include video metadata (views, likes), influencer characteristics (follower count), and sales metrics. We extracted 22,980 consumer comments and classified each for actionable and ideational inspiration using a text-mining approach based on the customer inspiration scale [6]. Comments expressing purchase intentions or trial motivations were coded as actionable; those expressing new perspectives or possibilities as ideational. The sample comprises 268 unique influencers.

Hashtag classification. Content analysis yielded two categories and seven lower-level types:

(I)Concert hashtags, including (1) specific name tags (48.7%) (2) product category tags (54.2%), and (7) celebrity tags (0.2%).

(II)Abstract hashtags, including (3) life scenario tags (76.9%), (4) function description tags (22.3%), (5) value-oriented tags (4.3%) and (6) festival marketing tags (1.2%). (Tag types numbered in descending order of proportion)

A construal hierarchy index weighted each tag type along a concrete-abstract continuum based on CLT. The results of the inspiration classification and other key information are presented in Table 1.

Table 1. Descriptive Statistics (N = 999)

Variable	M	SD	Min	Max
Act. Insp. (Resid.)	0.000	0.225	-0.330	0.903
Ide. Insp. (Resid.)	0.000	0.043	-0.016	0.989
ln(Sales)	2.724	2.201	0.000	11.379
ln(Views)	8.854	2.843	0.693	17.778
ln(Followers)	10.032	1.786	6.797	16.168
Hashtag Count	3.380	1.521	0.000	7.000
Prod. Cat. Tag (0/1)	0.542	0.499	0.000	1.000
Life Scen. Tag (0/1)	0.769	0.422	0.000	1.000
Hierarchy Index	-0.063	0.603	-1.000	1.000

Note. Resid. = residualized on ln(views) and ln(followers).

3.2 Measures and estimation

The dependent variables are residualized proportions of actionable and ideational inspiration, removing the mechanical influence of video exposure on comment composition. Controls include ln (followers) and ln (total comments). The main estimation employs OLS with brand and quarter fixed effects.

$$Insp_i = \beta_0 + \beta_1 Tag_1 i + \beta_2 Tag_2 i + \ldots + \beta_7 Tag_7 i + \gamma_1 \ln(Followers) + \gamma_2 \ln(Comments) + \mu_b + \tau_q + \varepsilon_i \quad (1)$$

where Tag_{ki} denotes the presence (1) or absence (0) of hashtag content type k in video i ($k = 1,\ldots,7$); μb and τq are brand and quarter fixed effects; εi is the error term. Standard errors are heteroskedasticity-consistent (HC1).

Supplementary analyses examine tag density moderation, inspiration complementarity, and downstream sales.

$$InspRatio_i = ActionableComments_i / TotalComments_i \quad (2)$$

$$ActInsp_i = \text{Residual}(InspRatio_i \mid \ln(Views_i), \ln(Followers_i)) \quad (3)$$

The results show a marginal U-shaped relationship, consistent with Jenders et al. [15,18].

4. Result

4.1 Effects of Hashtag Content Types

Table 2 presents OLS estimates. Product category tags significantly increase actionable inspiration ($\beta = .040$, SE = .017, $p = .022$), while the other six tag types are non-significant. For ideational inspiration, festival marketing tags ($\beta = -.008$, $p = .008$) and celebrity tags ($\beta = -.015$, $p < .001$) show negative effects, though these are based on very few videos (12 and 2, respectively). Product category tags do not affect ideational inspiration ($\beta = .004$, $p = .267$), confirming effect specificity. The models explain 13.5% of variance in actionable inspiration (Adj. $R^2 = .121$) but only 1.5% in ideational inspiration. Follower count positively predicts actionable inspiration ($\beta = .015$, $p < .001$).

Table 2. Main Regression Results

	M1	M2	M3	M4	M5	M6
DV	Act.Insp.	Ide.Insp.	Act.Insp.	Ide.Insp.	Act.Insp.	Ide.Insp.
Prod. Cat.	0.040* (0.017)	0.004 (0.004)	—	—	—	—
Life Scen.	-0.011 (0.018)	-0.005 (0.004)	—	—	—	—
Function	-0.005 (0.028)	-0.003 (0.006)	—	—	—	—
Spec. Name	0.016 (0.016)	0.000 (0.003)	—	—	—	—
Festival	0.039 (0.062)	-0.008** (0.003)	—	—	—	—
Celebrity	0.001 (0.152)	-0.015*** (0.004)	—	—	—	—
Hierarchy	—	—	-0.009 (0.013)	-0.003 (0.003)	—	—
Diversity	—	—	0.012 (0.009)	-0.001 (0.002)	—	—
Tag Count	—	—	—	—	-0.043 (0.028)	-0.007 (0.005)
Tag Count²	—	—	—	—	0.007† (0.004)	0.001 (0.001)
ln (Followers)	0.015*** (0.004)	0.000 (0.001)	0.015*** (0.004)	0.000 (0.001)	0.015*** (0.004)	0.000 (0.001)
ln (Comments)	-0.012† (0.007)	0.002 (0.001)	-0.012† (0.007)	0.002† (0.001)	-0.012† (0.007)	0.002 (0.001)
Brand FE	Yes	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes

	M1	M2	M3	M4	M5	M6
N	999	999	999	999	999	999
R^2	0.135	0.015	0.132	0.013	0.133	0.013
Adj. R^2	0.121	-0.002	0.122	0.002	0.123	0.002

4.2 Construal Level and Hashtag Count

The construal hierarchy index is non-significant for both inspiration types (M3: $\beta = -.009$, $p = .487$; M4: $\beta = -.003$, $p = .249$), contradicting CLT predictions. Tag diversity also fails to reach significance. Hashtag count exhibits a marginal U-shaped relationship with actionable inspiration (M5: quadratic $\beta = .007$, $p = .090$; minimum at ≈3 tags), but not with ideational inspiration ($p = .106$).

4.3 Supplementary Analyses

Tag density moderation. Life scenario tags become strongly positive in tag-dense videos (≥4 tags, N = 523: $\beta = .128$, $p < .001$), while product category tags lose significance ($\beta = .015$, $p = .506$). In low-density videos (N = 476), neither is significant. This suggests abstract tag content requires cognitive scaffolding from surrounding tags to translate into inspiration.

Complementarity. Actionable and ideational inspiration are positively correlated ($r = .68$ full sample; $r = .78$ co-occurrence subsample of 135 videos), supporting complementarity rather than substitutability.

Downstream sales. Bootstrap mediation analysis (5,000 resamples) reveals that, hashtags do not have a significant direct effect on sales ($\beta = .194$, $p = .540$); however, they still exert an indirect effect through action-oriented inspiration: functional description tags show a significant indirect effect (ab = .020, 95% CI [0.001, 0.038]), as do specific name tags (ab = .012, CI [0.002, 0.023]). In both pathways, the total effect of tags on sales is near zero while the indirect path is significantly positive, indicating full mediation—tags drive sales by inspiring consumers rather than through direct conversion. The direct effect of inspiration on sales is significant without brand FE ($\beta = 1.811$, $p < .001$) but attenuates with brand FE ($\beta = .194$, $p = .540$), and the indirect effect does not persist under creator FE (ab = −.0002, CI [−.002, .001]). Adding the seven tag types to a base model yields $\Delta R^2 = .005$, a modest increment given the extensive controls.

Figure 1 illustrates the moderating effect of label density on the label type effect.

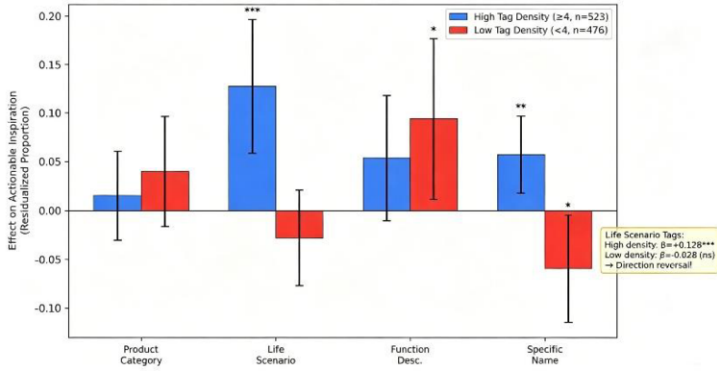


Figure 1. Tag Density Subgroup Comparison

Note: This figure is partially generated with the assistance of an AI image-generation model and manually revised by the authors.

5. Discussion

Theoretical implications. This study makes three contributions. First, we shift hashtag research from quantity to content type, showing that what brands hashtag matters more than how many. Product category tags uniquely drive actionable inspiration—no other tag type exhibits comparable effects—implying that hashtag effects are content-specific. Second, we extend customer inspiration theory to short-video marketing, documenting that actionable and ideational inspiration have distinct antecedents and are complementary ($r = .68–.78$) rather than substitutive, challenging the unitary construct assumption. Third, the null finding for CLT's construal hierarchy index is theoretically significant: the mechanism operates through information accessibility rather than construal level per se. Product category tags make action-related knowledge directly accessible [19], whereas abstract lifestyle tags require contextual scaffolding from co-occurring tags to become accessible. Information foraging theory [20], consistent with broader evidence that information quantity and quality jointly shape decision effectiveness [21] and that excessive information load can impair consumer choices [22], offers a compatible perspective: richer informational cues provide stronger "information scent," helping consumers link abstract content to specific needs. This explains why tag density moderates the effect of lifestyle tags but not product tags: abstract tags require surrounding cues to become cognitively accessible, whereas concrete tags are self-sufficient. Finally, the inspiration-to-sales link ($\beta = 1.811$, $p < .001$ without brand FE; $\beta = .194$, $p = .540$ with brand FE) provides initial evidence of inspiration as a commercial value driver, contributing to debates about whether customer engagement should prioritize delight [23, 24] or actionable relevance, though the attenuation under brand FE warrants caution.

Managerial implications. Brand managers should prioritize product category tags to drive behavioral engagement. For life scenario tags, a tag-stacking strategy—pairing them with three or more additional hashtags—activates their inspirational potential. Celebrity and festival tags, while valuable for reach, do not automatically generate inspirational engagement. The inspiration-sales link suggests monitoring actionable inspiration proportions in comments as a real-time diagnostic metric.

6. Limitation

Data Resource. Our data are from a single platform (Douyin) and industry (beauty), limiting generalizability. Experimental designs could strengthen causal claims. Text-mining-based inspiration measures may not fully capture psychometric nuance. The cross-sectional design does not capture temporal dynamics. The null CLT finding requires further investigation with alternative operationalizations (e.g., LIWC-based indices). Future research should examine brand-level moderators of the inspiration-sales link and test the framework across diverse platforms and product categories. Methodology. OLS with fixed effects cannot rule out reverse causality (e.g., trending products attracting more hashtags) or omitted variable bias (e.g., brand marketing budgets). Additionally, the tag density moderation was identified through exploratory analysis rather than a priori hypothesis. Although the pattern is consistent with construal-level theory—abstract tags require contextual scaffolding from co-occurring tags—the proposed mechanism (information accessibility) is not directly tested. Future research could manipulate tag density experimentally to provide more direct evidence.

References

[1] Toni M, Mattia G. TikTok hashtag challenge: a new digital strategy for consumer brand engagement (CBE). In: Proceedings of the European Marketing Academy (EMAC) 50th Annual Conference; 2021. Paper ID 104352.

[2] Cao Y, Lin J, Zhou Z. Promoting customer value co-creation through social capital in online brand communities: the mediating role of member inspiration. Comput Hum Behav. 2022;137:107440.

[3] Herhausen D, Kleijnen K, Verleye K, Verlegh P. Multichannel customer inspiration. J Interact Mark. 2019;47:47-64.

[4] Trope Y, Liberman N. Construal-level theory of psychological distance. Psychol Rev. 2010;117(2):440-63.

[5] Collins, A. M., & Loftus, E. F. A spreading-activation theory of semantic processing. Psychological Review.1975; 82(6):407–428.

[6] Böttger T, Rudolph T, Evanschitzky H, Pfrang T. Customer inspiration: conceptualization, scale development, and validation. J Mark. 2017 Nov;81(6):116-31.

[7] Thrash TM, Elliot AJ. Inspiration as a psychological construct. J Pers Soc Psychol. 2003;84(4):871-89.

[8] Ki CW, Park S, Kim YK. Investigating the mechanism through which consumers are "inspired by" social media influencers and "inspired to" adopt influencers' exemplars as social defaults. J Bus Res. 2022;144:264-77.

[9] Izogo EE, Mpinganjira M. Behavioral consequences of customer inspiration: the role of social media inspirational content and cultural orientation. J Res Interact Mark. 2020;14(4):431-59.

[10] McColl-Kennedy JR, Zaki M, Lemon KN, Urmetzer F, Neely A. Gaining customer experience insights that matter. J Serv Res. 2019;22(1):8-26.

[11] Miller GA. The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol Rev. 1956;63(2):81-92.

[12] Nam H, Joshi YV, Kannan PK. Harvesting brand information from social tags. J Mark. 2017;81(4):88-108.

[13] Nam H, Kannan PK. The informational value of social tagging networks. J Mark. 2014;78(4):21-40.

[14] Shin J, Chae H, Ko E. The power of e-WOM using the hashtag: focusing on SNS advertising of SPA brands. Int J Advert. 2018;37(4):1-15.

[15] Jenders M, Kacpzyk G, Kanhard H. Analyzing the viral spread of hashtags. In: Proceedings of the 24th ACM Conference on Hypertext and Social Media. New York: ACM; 2013. p. 35-40.

[16] Jin HS, Kim N. Can too many hashtags backfire? Understanding audiences' perceptions of hashtags on Instagram. Int J Advert. 2026:1-30.

[17] Ye N, Hashim NH, Baghirov F, Murphy J. Gender differences in Instagram hashtag use. J Hosp Mark Manag. 2018;27(4):386-404.

[18] Pierce JR, Aguinis H. The too-much-of-a-good-thing effect in management. J Manag. 2013;39(2):313-38.

[19] Higgins, E. T. Knowledge activation: Accessibility, applicability, and salience. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles*. 1996;133–168.

[20] Pirolli, P., & Card, S. Information foraging. Psychological Bulletin. 1999; 127(6): 833–851.

[21] Keller KL, Staelin R. Effects of quality and quantity of information on decision effectiveness. J Consum Res. 1987;14(2):200-13.

[22] Malhotra NK. Information load and consumer decision making. J Consum Res. 1982;8(4):419-30.

[23] Dixon M, Freeman K, Toman N. Stop trying to delight your customers. Harv Bus Rev. 2010 Jul;88(7/8):1-6.

[24] Rust RT, Oliver RL. Should we delight the customer? J Acad Mark Sci. 2000;28(1):86-94.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261091

Research on the Impact of Digital-Intelligent Transformation on Corporate Performance of Manufacturing Enterprises

Lingling Hou[a] and Jianhui Tan[b,1]
[a] *School of Economics and Management, Xiangnan University, Chenzhou, China*
[b] *Chenzhou Vocational Technical College, Chenzhou, China*

Abstract. Based on panel data of A-share listed manufacturing firms from 2010 to 2024, this paper constructs a two-way fixed-effects model to explore how digital-intelligent transformation affects corporate financial performance. Taking total asset turnover and R&D innovation investment as dual mediating variables, this study further conducts regional heterogeneity analysis. The empirical results show that digital-intelligent transformation significantly improves corporate ROE, which passes multiple robustness tests. Dual partial mediating effects are verified: digital-intelligent transformation boosts profitability by optimizing asset operation efficiency and stimulating R&D input. Heterogeneity analysis reveals that the performance promotion effect is prominent in eastern China's manufacturers yet insignificant in central and western regions. This paper enriches the internal mechanism theory of digital-intelligent transformation and provides differentiated policy and operational suggestions for governments and manufacturing enterprises across regions. The novelty of this study lies in integrating asset operation and R&D innovation into a unified mediating framework to reveal the dual-path mechanism, and in revealing the spatial unevenness of transformation dividends across China's regions, providing micro-level empirical evidence for differentiated manufacturing digitalization policies.

Keywords. Digital-Intelligent Transformation, Corporate Performance, Manufacturing Listed Firms, Mediating Effect, Regional Heterogeneity

1.Introduction

With the continuous deepening of a new round of global technological revolution and industrial transformation, digital technologies such as the Internet of Things, big data and artificial intelligence have deeply penetrated the real economy, driving the comprehensive iteration and upgrading of global manufacturing production and organizational models toward digitalization and intelligence. China has successively

[1]Corresponding Author: Jianhui Tan, Chenzhou Vocational Technical College, Chenzhou, China; E-mail: 123722826@qq.com.

This paper is supported by ①Project of Hunan Social Science Achievement Evaluation Committee (No. XSP25YBC486); ②Hunan Provincial Teaching Reform Research Project (No. 202502001402); ③Hunan Provincial College Students' Innovation Training Program (Document No. Xiangjiaotong [2025] 141- Item No.3109).

launched top-level strategies including "Manufacturing Power" and "Digital China", and continuously strengthened policy support for the digital-intelligent upgrading of manufacturing industry, aiming to rely on digital technological changes to resolve structural development bottlenecks of traditional manufacturing sectors such as high costs, low production and operation efficiency, and lagging response to market demand. Against this backdrop, systematically analyzing the internal mechanism through which digital-intelligent transformation affects the performance of manufacturing enterprises carries important theoretical value and industrial practical significance.

At present, extensive research has been conducted by domestic and foreign academic circles on the correlation between digital-intelligent transformation and corporate performance, yet divergences and research gaps still exist in research conclusions and analytical perspectives. Longfei M et al. (2024) argued that digital-intelligent transformation can reduce operating costs and improve financial performance of manufacturing enterprises, without distinguishing short-term and long-term transformation effects [1]. Rahman A S et al. (2025) proposed that digital perception, digital capture and digital conversion capabilities are core driving factors of digital-intelligent transformation, while transformation costs will weaken the positive impact of transformation on performance [2]. Taking A-share manufacturing enterprises in Shanghai and Shenzhen Stock Exchanges as samples, Zhou and Li (2025) found that digital transformation can improve corporate performance through multiple channels including management efficiency improvement, cost reduction and innovation upgrading [3]. Some literature has drawn differentiated research conclusions: targeting non-high-tech manufacturing industries, Qi and Cai (2020) found that the impacts of digitalization on management and sales activities offset each other, ultimately leading to no significant change in comprehensive corporate performance [4]. Liang et al. (2026) demonstrated that digital intelligence transformation significantly boosts firm performance, with stronger effects in state-owned enterprises and high-tech manufacturing sectors [5].

In terms of transmission mechanism research, most existing literature only tests a single mediating path separately, failing to incorporate asset operation capacity and R&D innovation capacity into a unified analytical framework for joint testing. Taking manufacturing enterprises in Guangdong Province as research samples, Zhu et al. (2023) verified the mediating role of technological innovation between digital-intelligent transformation and corporate performance [6]. Ma and Jaafar (2025) confirmed that resource allocation exerts a direct influence on short-term firm performance, while innovation capability plays a critical role in driving long-term performance [7]. Jiang et al. (2026) conducted a systematic review and emphasized that the relationship between digital transformation and sustainable firm performance is context-dependent, requiring careful consideration of industry and regional factors [8]. Meanwhile, the sample time span of relevant existing literature is generally short, which cannot fully capture the long-term cumulative operational performance effects generated by digital-intelligent transformation.

In view of the above research gaps, this paper takes A-share listed manufacturing companies in China from 2010 to 2024 as a long-cycle research sample, constructs a two-way fixed-effects model to systematically investigate the direct impact of digital-intelligent transformation on corporate performance, tests the dual mediating transmission mechanisms of total asset turnover ratio and innovation investment, and conducts heterogeneity analysis by dividing samples according to geographical locations of eastern, central and western regions. Specifically, this study addresses three

research questions: (1) Does digital-intelligent transformation significantly improve the financial performance of manufacturing firms? (2) Through which mechanisms—asset operation optimization or R&D innovation stimulation—does transformation affect performance? (3) Does the performance-enhancing effect differ across China's eastern, central, and western regions? The marginal contributions of this paper are mainly reflected in two aspects: first, it integrates two core transmission paths of asset operation and R&D innovation, improves the mechanism analysis system through which digital-intelligent transformation empowers corporate performance, and makes up for the limitation of single mediating test in existing literature; second, it analyzes the spatial differentiation of transformation effects combined with the unbalanced regional development characteristics of China, providing decision-making basis for differentiated and coordinated promotion of digital-intelligent upgrading of manufacturing industry nationwide.

2.Theoretical Analysis and Research Hypotheses

2.1 Digital-Intelligent Transformation and Corporate Performance

Based on the resource-based view, data assets, digital technologies, intelligent production systems and other elements accumulated in the process of digital-intelligent transformation belong to scarce and inimitable heterogeneous resources of enterprises [9]. Relying on digital technologies, enterprises integrate production factors including human resources, materials and capital, optimize resource allocation structure, reduce resource waste, and effectively cut operating costs. From the perspective of dynamic capability theory, big data and IoT technologies help enterprises capture real-time market demands and predict industry trends, prompting enterprises to rapidly adjust business strategies and enhance organizational flexibility and market response capabilities [10]. Meanwhile, digital-intelligent transformation drives the flattening reform of corporate organizational structures, compresses management levels and improves decision-making efficiency; intelligent production lines and automated equipment replace repetitive manual procedures, continuously releasing efficiency dividends at the production end. Under the framework of Industry 4.0, cyber-physical systems, digital twins and other technologies reconstruct traditional production modes to achieve cost reduction and efficiency improvement in an all-round manner [11]. Superimposed by multiple effects, corporate profitability keeps improving. Accordingly, the research hypothesis is proposed as follows:

H1: Digital-intelligent transformation of manufacturing enterprises exerts a significant positive promoting effect on corporate performance.

2.2 Mediating Effect of Total Asset Turnover Ratio

Combined with the Industry 4.0 framework and dynamic capability theory, digital technologies realize full-process dynamic digital management and control covering production and manufacturing, inventory management, supply chain collaboration and fixed asset operation and maintenance. IoT terminals collect real-time production and inventory data, and big data algorithms conduct accurate demand forecasting, effectively alleviating inventory backlog and long-term idle fixed assets, accelerating

inventory turnover and capital recovery efficiency; intelligent production scheduling systems and cross-enterprise supply chain collaboration platforms optimize the whole-link operation process, and systematically improve the comprehensive utilization efficiency of stock assets [12].

Total asset turnover ratio is a core financial indicator measuring corporate asset management level. The higher the asset turnover efficiency, the stronger the capacity of enterprises to convert stock assets into operating income, the higher the operating income created by unit assets, and the improved corporate profitability accordingly. It can be deduced that digital-intelligent transformation can improve corporate financial performance by increasing total asset turnover ratio. Therefore, the hypothesis is put forward:

H2: Total asset turnover ratio plays a positive mediating transmission effect between digital-intelligent transformation and corporate performance.

2.3 Mediating Effect of Innovation Investment

Digital-intelligent transformation provides underlying technical support and digital tool empowerment for corporate R&D and innovation activities. On the one hand, digital R&D simulation platforms, virtual testing, digital twins and other technical means can shorten the R&D cycle of new products, reduce innovation trial-and-error costs, strengthen enterprises' willingness to increase R&D investment, and drive enterprises to continuously expand investment in R&D funds, R&D manpower and other innovation resources; on the other hand, digital-intelligent transformation builds an open and collaborative industrial ecosystem, facilitating enterprises to integrate internal and external innovation resources and expand the boundary of technological R&D.

Sustained increased innovation investment will drive enterprises to realize technological iteration and differentiated product upgrading, build unique market competitive advantages, raise product added value and industry bargaining power, and ultimately translate into growth of corporate operating performance. Yin et al. (2022) also confirmed that digital-intelligent transformation can indirectly improve financial performance by enhancing corporate innovation capacity [13]. In short, digital-intelligent transformation can drive enterprises to expand the scale of innovation investment, and realize performance growth relying on improved innovation capacity. Accordingly, the hypothesis is proposed:

H3: Innovation investment plays a positive mediating transmission effect between digital-intelligent transformation and corporate performance.

3.Research Design

3.1 Sample Selection and Data Sources

This paper takes A-share listed manufacturing companies in China from 2010 to 2024 as the initial research sample. The sample starts from 2010 when China's "Twelfth Five-Year Plan" identified intelligent manufacturing as a strategic priority. The data are processed through the following screening criteria: (1) Eliminate observation samples of enterprises with abnormal financial status and ongoing operation risks such as ST and *ST; (2) Eliminate samples with missing values of core variables; (3) Perform 1% winsorization on all continuous variables to eliminate interference of extreme values.

Finally, 24,126 valid observations are obtained. The data comes from two sources: firm-level financial and governance data from the CSMAR database, and digital-intelligent transformation keyword frequencies extracted from annual reports of listed companies using Python-based text analysis. Data processing and statistical tests are completed with Stata software.

3.2 Variable Definition

3.2.1 Explained Variable

The explained variable refers to corporate performance. Existing literature divides measurement indicators of corporate performance into financial performance and non-financial performance indicators. Financial indicators have the advantages of data availability and quantitative objectivity, and are most widely applied in relevant empirical research. This paper adopts Return on Equity (*ROE*, the ratio of net profit to owners' equity) as the core measurement indicator of corporate performance. In addition, Return on Assets (*ROA*, the ratio of net profit to average total assets) is introduced for subsequent robustness tests to alleviate measurement bias caused by a single indicator.

3.2.2 Explanatory Variable

The explanatory variable is the digital-intelligent transformation level of manufacturing enterprises (*lnDig*). For its measurement, text analysis is widely adopted in academic circles to sort out word frequency data of digital-intelligent transformation in annual reports of listed companies to measure the transformation level [14][15]. Referring to the research of Zhao et al. [14], this paper defines the connotation of digital-intelligent transformation from four dimensions: digital technology application, intelligent manufacturing, Internet business model and modern information system, and establishes the corresponding keyword database. Python technology is used to read keyword information of digital-intelligent transformation in annual reports of listed companies respectively, count the total word frequency, add 1 to the total word frequency and take natural logarithm. This design can objectively reflect the implementation intensity of corporate digital-intelligent transformation and avoid regression bias brought by zero-value samples, so as to characterize the digital-intelligent transformation level of enterprises.

3.2.3 Mediating Variables

To clarify the two internal transmission paths through which digital-intelligent transformation acts on corporate performance, two mediating variables are set in this paper: (1) Total Asset Turnover (*ATO*), the ratio of operating income to average total assets, comprehensively reflecting the operation and circulation efficiency of all stock assets of enterprises. (2) Innovation Investment (*Rdsz*), the ratio of annual total R&D investment to average total assets, used to measure the investment intensity of corporate R&D resources.

3.2.4 Control Variables

Referring to the research of Shang et al. (2024) [16] and Qian et al. (2018) [17], this paper selects the following 8 control variables to reduce their interference on the research: (1) Enterprise Size (*Size*), reflecting the business volume and resource

acquisition capacity of enterprises; (2) Ownership Concentration (*Top1*), embodying the corporate ownership governance structure and decision-making dominant power; (3) Firm Age (*Age*), characterizing the operation duration of enterprises; (4) Management Expense Ratio (*Mfee*), reflecting the internal management cost control level; (5) Dual Leadership (*Dual*), indicating whether the chairman concurrently serves as the general manager, reflecting the corporate governance structure; (6) Asset-Liability Ratio (*Lev*), measuring the financial leverage level of enterprises; (7) Audit Opinion (*Opinion*); (8) Independent Director Ratio (*Indep*).

Table 1. Variable Definition

Variable Type	Variable Name	Symbol	Variable definition
Explained Variable	Corporate Performance	*ROE*	Return on Equity (Net Profit / Owners' Equity)
Explanatory Variable	Digital-Intelligent Transformation	*lnDig*	ln(Word Frequency of Digital-Intelligent Keywords + 1)
Mediating Variable	Total Asset Turnover	*ATO*	Operating Income / Average Total Assets
	Innovation Investment	*Rdsz*	Total R&D Investment / Average Total Assets
Control Variable	Enterprise Size	*Size*	ln(Total Year-End Assets)
	Ownership Concentration	*Top1*	Shareholding Ratio of the Largest Shareholder
	Firm Age	*Age*	ln(Current Year - Establishment Year + 1)
	Management Expense Ratio	*Mfee*	Management Expense / Operating Income
	Dual Leadership	*Dual*	1 if the chairman concurrently serves as general manager, otherwise 0
	Asset-Liability Ratio	*Lev*	Total Year-End Liabilities / Total Year-End Assets
	Audit Opinion	*Opinion*	1 for standard unqualified audit opinion, otherwise 0
	Independent Director Ratio	*Indep*	Number of Independent Directors / Total Number of Board Directors

3.3 Model Construction

Hausman test is conducted first, and the p-value of the test result is 0.0000, which rejects the null hypothesis of random effects model at the 1% significance level. Therefore, the fixed-effects model is adopted.

To verify the main effect Hypothesis H1, the model is constructed as follows:

$$ROE_{i,t}=\alpha_0+\alpha_1 ln\,Dig_{i,t}+\sum Control_{i,t}+\sum Year+\sum Id+\varepsilon_{i,t} \quad (1)$$

To verify Hypothesis H2, the following models are constructed:

$$ATO_{i,t}=\alpha_0+\alpha_1 ln\,Dig_{i,t}+\sum Control_{i,t}+\sum Year+\sum Id+\varepsilon_{i,t} \quad (2)$$

$$ROE_{i,t}=\alpha_0+\alpha_1 ln\,Dig_{i,t}+\alpha_2 ATO_{i,t}+\sum Control_{i,t}+\sum Year+\sum Id+\varepsilon_{i,t} \quad (3)$$

To verify Hypothesis H3, the following models are constructed:

$$Rdsz_{i,t}=\alpha_0+\alpha_1 ln\,Dig_{i,t}+\sum Control_{i,t}+\sum Year+\sum Id+\varepsilon_{i,t} \quad (4)$$

$$ROE_{i,t}=\alpha_0+\alpha_1 ln\,Dig_{i,t}+\alpha_2 Rdsz_{i,t}+\sum Control_{i,t}+\sum Year+\sum Id+\varepsilon_{i,t} \quad (5)$$

In the above research model, *ROE* is the explained variable; *lnDig* is the explanatory variable; *Control* represents a series of control variables, *ATO* and *Rdsz* serve as the mediating variables. *Year* and *Id* are the year fixed effect and firm individual fixed effect, ε stands for the random error term, and subscripts *i* and *t* correspond to the firm dimension and time dimension.

4 Empirical Analysis

4.1 Descriptive Statistics

Table 2 reports the descriptive statistical results of main variables. The average Return on Equity (*ROE*) of sample enterprises is 0.0560, ranging from -0.644 to 0.311 with a standard deviation of 0.125, indicating significant differentiation in internal profitability of listed manufacturing companies. The average value of the core explanatory variable digital-intelligent transformation level (*lnDig*) is 2.941, with a maximum value of 5.670, a minimum value of 0 and a standard deviation of 1.147, reflecting that the promotion degree of digital-intelligent transformation of manufacturing enterprises varies greatly, and the unbalanced characteristic of digital development is prominent. The statistical distribution characteristics of other mediating variables and control variables are consistent with the conclusions of existing mainstream literature, proving the rationality of sample data.

Table 2. Descriptive statistical analysis

Variables	N	Mean	Median	SD	Min	Max
ROE	24126	0.0560	0.0660	0.125	-0.644	0.311
lnDig	24126	2.941	2.944	1.147	0	5.670
ATO	24126	0.620	0.551	0.342	0.118	2.106
Rdsz	24126	0.0240	0.0210	0.0180	0	0.0970
Size	24126	22.04	21.89	1.142	19.97	25.52
Top1	24126	0.333	0.311	0.138	0.0920	0.715
Age	24126	2.927	2.944	0.318	1.946	3.584
Mfee	24126	0.0810	0.0690	0.0550	0.0100	0.332
Dual	24126	0.329	0	0.470	0	1
Lev	24126	0.391	0.383	0.190	0.0530	0.854
Opinion	24126	0.977	1	0.150	0	1
Indep	24126	0.376	0.333	0.0530	0.333	0.571

4.2 Benchmark Regression Analysis

To avoid the interference of multicollinearity on the effectiveness of model estimation, Variance Inflation Factor (VIF) is adopted for collinearity diagnosis in this paper. After inspection, the VIF values of all variables are less than 2, far below the empirical threshold of 10. It shows that there is no serious multicollinearity among variables, which can effectively guarantee the reliability and validity of subsequent regression analysis.

Furthermore, benchmark regression analysis is carried out, and the results are shown in Table 3. Column (1) does not control individual and year fixed effects and excludes control variables; the regression coefficient of digital-intelligent transformation level is 0.003, significant at the 1% confidence level. Column (2) incorporates two-way fixed effects of individual enterprises and years, and the regression coefficient of lnDig is 0.008, significant at the 1% confidence level. Column (3) is the complete model with both fixed effects and control variables added, and the regression coefficient of lnDig is 0.005, also significant at the 1% confidence level.

Therefore, whether the model contains control variables or fixed effects, the coefficient of digital-intelligent transformation (*lnDig*) of manufacturing enterprises is significantly positive. This empirical result strongly confirms that promoting digital-intelligent transformation can significantly improve corporate performance, thus verifying Hypothesis H1 proposed in this research.

Table 3. Benchmark Regression Results

	(1)	(2)	(3)
	ROE	*ROE*	*ROE*
lnDig	0.003***(0.001)	0.008***(0.001)	0.005***(0.001)
Constant	0.049***(0.002)	0.034***(0.004)	-0.495***(0.057)
Control Variables	NO	NO	YES
Id	NO	YES	YES
Year	NO	YES	YES
N	24126	23772	23772
R^2	0.001	0.392	0.515
adj.R^2	0.001	0.298	0.439

Notes: *, **, and *** denote significance levels at 10%, 5%, and 1%, respectively. Standard errors are reported in parentheses. The same applies hereinafter.

4.3 Robustness Test

To verify the stability of benchmark regression conclusions, robustness tests are carried out from two dimensions: variable replacement and sample interval adjustment.

(1) Replacing the Explained Variable Considering that the measurement method of corporate performance may affect the results, Return on Assets (*ROA*) is used to replace Return on Equity (*ROE*) as the explained variable for robustness test. The results are shown in Column (1) of Table 4. The coefficient of lnDig is significantly positive at the 1% level, indicating that the research conclusion remains robust under different measurement methods of corporate performance.

(2) Adjusting Sample Scope Considering that time selection may interfere with research results, the sample time range is shortened to 2015–2024 for regression analysis. The results are shown in Column (2) of Table 4; the positive impact of digital-intelligent transformation on corporate performance is still significantly positive at the 1% significance level, proving that the conclusion holds after changing the sample interval.

Table 4. Robustness Test Results

	(1)Replace Explained Variable	(2)Adjust Sample Interval
	ROA	*ROE*
lnDig	0.001***(0.000)	0.006***(0.002)
Constant	-0.058**(0.024)	-0.768***(0.091)
Control Variables	YES	YES
Id	YES	YES
Year	YES	YES
N	23772	17865
R^2	0.627	0.570
adj.R^2	0.569	0.481

4.4 Mediating Mechanism Test

Table 5 reports the mediating test results of total asset turnover ratio and innovation investment. Column (1) regresses the explanatory variable lnDig on the mediating variable *ATO*; the coefficient of lnDig is 0.008 and significantly positive at the 1% level, indicating that digital-intelligent transformation can effectively improve corporate asset turnover efficiency. Column (2) incorporates both *lnDig* and *ATO* to regress *ROE*, and the coefficients of the two variables are significantly positive at the 1% level, which confirms that total asset turnover ratio exerts a partial mediating effect, and Hypothesis H2 is verified.

Column (3) regresses the explanatory variable *lnDig* on the mediating variable *Rdsz*; the coefficient of *lnDig* is 0.001 and significantly positive at the 1% level, illustrating that digital-intelligent transformation can effectively encourage enterprises to expand R&D investment. Column (4) incorporates both *lnDig* and *Rdsz* to regress *ROE*, and the two coefficients are significantly positive at the 1% and 5% levels respectively, proving that innovation investment also undertakes a partial mediating transmission role, and Hypothesis H3 is verified.

In summary, the digital-intelligent transformation of manufacturing enterprises has both a direct promoting effect on corporate performance, and two indirect transmission paths of increasing total asset turnover ratio and expanding innovation investment.

Table 5. Mediating Mechanism Test Results

	(1) *ATO*	(2) *ROE*	(3) *Rdsz*	(4)*ROE*
lnDig	0.008***(0.002)	0.004***(0.001)	0.001***(0.000)	0.004***(0.001)
ATO		0.103***(0.004)		
Rdsz				0.177**(0.081)
Constant	2.770***(0.090)	-0.780***(0.057)	0.090***(0.005)	-0.511***(0.057)
Control Variables	YES	YES	YES	YES
Id	YES	YES	YES	YES
Year	YES	YES	YES	YES
N	23772	23772	23772	23772
R^2	0.837	0.528	0.827	0.515
adj.R^2	0.812	0.454	0.800	0.439

4.5 Heterogeneity Analysis

There are significant endowment gaps in digital infrastructure, industrial clusters and talent factors among eastern, central and western regions of China, which directly lead to spatial differentiation characteristics of performance-driven effects brought by digital-intelligent transformation. The eastern region has sound supporting systems including industrial Internet, 5G and cloud computing, high industrial chain collaboration degree and prominent spillover effect of digital technologies; the central and western regions are faced with prominent shortcomings in digital infrastructure, scattered manufacturing layout, high digital transformation costs and low technology absorption efficiency, making it difficult to quickly realize transformation benefits. To quantitatively verify such spatial differences, this paper divides samples into three sub-groups of eastern, central and western regions according to geographical locations for grouped regression.

The grouped regression results are shown in Table 6. The coefficient of *lnDig* in the eastern region is 0.004, significantly positive at the 1% level; although the coefficients of *lnDig* in central and western regions are positive, they fail to pass the significance test. This result indicates that the promoting effect of digital-intelligent transformation on corporate performance is significantly established for manufacturing enterprises in the eastern region, while digital transformation dividends have not been fully released in central and western regions. The eastern region relies on a complete digital industrial ecosystem to form a positive cycle mechanism where digital input drives performance growth, while central and western regions are constrained by resource endowments, limiting the empowerment effect of digital technologies.

Table 6. Heterogeneity Analysis

	(1)Eastern Region	(2)Central Region	(3)Western Region
	ROE	*ROE*	*ROE*
lnDig	0.004*** (0.001)	0.004 (0.004)	0.005 (0.003)
Constant	-0.567*** (0.065)	-0.037 (0.188)	-0.608*** (0.144)
Control Variables	YES	YES	YES
Id	YES	YES	YES
Year	YES	YES	YES
N	16815	2975	3977
R^2	0.525	0.504	0.517
adj.R^2	0.447	0.430	0.442

5 Research Conclusions, Discussion and Policy Suggestions

5.1 Research Conclusions and Discussion

Based on the panel data of A-share listed manufacturing enterprises in China from 2010 to 2024, this paper empirically investigates the impacts, transmission mechanisms and regional heterogeneity of digital-intelligent transformation on manufacturing firm performance. The results show that: First, digital-intelligent transformation significantly improves firm performance, and this finding remains valid after a series of robustness tests. Second, total asset turnover and R&D innovation investment exert

partial mediating effects. Digital-intelligent transformation can boost firm performance directly, and also indirectly enhance performance by optimizing asset operation efficiency and strengthening innovation capacity. Third, the transformation effect presents obvious regional heterogeneity. Digital-intelligent transformation achieves remarkable outcomes among eastern manufacturing firms, while the transformation dividends in central and western regions have not been fully released, leading to unbalanced regional development.

The positive direct effect is consistent with recent findings by Longfei et al. (2024), Zhou and Li (2025), and Liang et al. (2026) [1][3][5]. This study extends prior work by empirically distinguishing two partial mediating channels: Total Asset Turnover and Innovation Investment. The results confirm that both operational efficiency improvement and innovation upgrading serve as transmission mechanisms—a dual-path distinction absent in prior single-path analyses. This finding complements Liang et al. (2026), who identified cost reduction and innovation enhancement as dual indirect paths [5], and extends Yin et al. (2022) by showing that innovation capacity is not the sole mediating mechanism [13]. The significant regional heterogeneity aligns with Jiang et al. (2026), who emphasized the context-dependence of digital returns [8], and echoes Qi and Cai (2020) regarding spatial imbalances in transformation outcomes [4].

5.2 Policy Suggestions

Corresponding suggestions are put forward from the perspectives of government and enterprises according to the research conclusions.

(1) Government perspective. Implement differentiated support policies to narrow regional gaps in transformation. Relying on digital infrastructure and industrial clusters, eastern regions should build demonstration platforms for intelligent manufacturing to facilitate the spillover of technologies and experience. Central and western regions need to improve new infrastructure, and reduce enterprises' transformation costs through subsidies, tax incentives and talent training. A coordination platform linking eastern and western regions should be constructed to promote cross-regional factor flow and realize the balanced development of manufacturing digital-intelligent transformation nationwide.

(2) Enterprise perspective. Promote digital-intelligent transformation in an orderly manner based on the two mediating paths. Firms should advance transformation in light of their scale, industrial characteristics and regional endowments and avoid blind heavy-asset investment. Internet of Things, big data and digital twin technologies can be adopted to revitalize stock assets and raise total asset turnover. Enterprises should establish regular R&D systems and leverage digital technologies to reduce innovation trial-and-error costs. Eastern firms can focus on high-end applications and model innovation; manufacturing firms in central and western regions should prioritize low-cost, lightweight transformation solutions and carry out reforms phase by phase.

5.3 Limitations and Future Research

This study has three main limitations. First, the analysis is confined to Chinese listed manufacturing firms, limiting generalizability to unlisted firms or other institutional contexts. Second, digital-intelligent transformation is measured by keyword frequency in annual reports, which captures rhetoric rather than actual implementation depth.

Third, the 15-year sample period covers significant economic fluctuations including the COVID-19 pandemic, which may have affected digital investment and performance unevenly. Future research could extend to cross-country comparisons, adopt alternative measures such as patent classifications in AI and IoT, and investigate AI-specific capabilities including AI-enabled decision-making, predictive maintenance, and intelligent logistics as distinct dimensions of digital transformation.

References

[1] Longfei M, Thanaphon R, Gumpora S. Research on the Impact of Digital Transformation on the Performance of Manufacturing Enterprises. Proceedings of Business and Economic Studies,2024,7(4):16-22. DOI:10.26689/PBES.V7I4. 8135.

[2] Rahman A S, Taghizadeh S N ,Tipu A A S, et al. Leveraging dynamic capabilities for digital transformation: Exploring the moderating role of cost in environmental performance of SMEs. Journal of Open Innovation: Technology, Market, and Complexity, 2025,11(2):100523. DOI:10.1016/J.JOITMC.2025.100523.

[3] Zhou Xiaoliang, Li Huipeng. The Impact of Digital Transformation in Manufacturing on Enterprise Performance. Shandong Social Sciences, 2025, (1):127-135.

[4] Qi Yudong, Cai Chengwei. Research on Multiple Impacts and Mechanisms of Digitalization on Manufacturing Enterprises' Performance. Study & Exploration, 2020, (7): 108-119.

[5] Liang C Y, Xu C F, Zhu K Y, et al. The impact and mechanism analysis of digital intelligence transformation on manufacturing enterprises' performance. Technology Analysis & Strategic Management, 2026. DOI:10.1080/09537325.2026.2660353.

[6] Zhu Jingting, Guo Feiyun, Lin Yun. Digitalization Level, Technological Innovation and Corporate Performance. The Chinese Certified Public Accountant, 2023, (9): 63-67.

[7] Ma G X, Jaafar H. Digital transformation and firm performance among Chinese manufacturing firms: A moderated-mediation model. Asian Journal of Economic Modelling, 2025, 13(4): 619-638. DOI:10.55493/5009.v13i4.5720

[8] Jiang C, Pimdee P, Sukkamart A. Digital transformation and sustainable firm performance in global manufacturing: A systematic review. Edelweiss Applied Science and Technology, 2026, 10(3): 277-297.DOI: 10.55214/2576-8484.v10i3.12343

[9] Nambisan S ,Wright M ,Feldman M. The digital transformation of innovation and entrepreneurship: Progress, challenges and key themes. Research Policy,2019,48(8):103773. DOI:10.1016/j.respol. 2019.03.018.

[10] Teece J D . Dynamic capabilities as (workable) management systems theory. Journal of Management & Organization, 2018, 24(3):359-368. DOI:10.1017/jmo.2017.75.

[11] André H ,René B ,David M , et al. A Systematic Review of the Literature on Digital Transformation: Insights and Implications for Strategy and Organizational Change. Journal of Management Studies,2020,58(5):1159-1197. DOI:10.1111/JOMS.12639.

[12] Wu Jie, Zhou Shengdi, Wang Gaiyin. Research on the Impact of Digital Transformation on Corporate Performance:An Empirical Analysis from the Perspective of Digital Depth and Breadth. Shanghai Journal of Economics, 2024, (9):49-59.

[13] YIN Xianan, ZHAN Ximing, TANG Shaoqing. The Impact Mechanism of Manufacturing Enterprises' Digital Transformation on Financial Performance. China Business and Market,2022,36(7):96-106.

[14] Zhao Chenyu, Wang Wenchun, Li Xuesong. How Does Digital Transformation Affect the Total Factor Productivity of Enterprises? Finance & Trade Economics,2021,42(7):114-129.

[15] Wu Fei, Hu Huizhi, Lin Huiyan, et al. Enterprise Digital Transformation and Capital Market Performance: Empirical Evidence from Stock Liquidity. Journal of Management World, 2021,37(7):130-144+10.

[16] Shang Hangbiao, Liu Jiaqi, Wang Zhilin, et al. Research on the Impact of Digital Transformation Differentiation on Enterprise Performance. Chinese Journal of Management, 2024, 21(2): 193-201. DOI:10.3969/j.issn.1672-884x. 2024.02.004.

[17] Qian Xuesong, Kang Jin, Tang Yinglun, et al. Industrial Policy, Efficiency of Capital Allocation and Firm's Total Factor Productivity—Evidence from a Natural Experiment in China. China Industrial Economics, 2018, (8):42-59.

Section 2

Track on “Applied-Education and Applied-Research in AI, Data Science, and Web6.0”, the Second Edition

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)

© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261093

The Algorithmic Co-Thinker: A Generative AI Framework for Persona-Aware Computational Mathematics

Allan M Canonigo[1]
National Institute for Science and Mathematics Education Development, University of the Philippines, Diliman, Quezon City 1101, Philippines
ORCiD ID: Allan M. Canonigo https://orcid.org/0000-0003-4330-1068

Abstract. The integration of Computational Thinking (CT) into mathematics education frequently encounters challenges within Digital Mathematics Learning Environments (DMLEs), which often adopt static, one-size-fits-all approaches that fail to adequately support learner diversity. While prior studies have identified distinct learner personas—such as the Frustrated Perfectionist and the Conceptually Stalled Learner—a critical barrier remains in scaling personalized, real-time pedagogical support. This study proposes a novel framework utilizing Generative AI (GenAI) to transform DMLEs into dynamic "co-thinkers." Employing a two-phase design-based research methodology, we first empirically validated a four-persona typology through qualitative analysis and quantitative behavioral signatures. Subsequently, we developed and expert-validated a conceptual GenAI scaffolding prototype, comprising structured prompt templates engineered to deliver metacognitive and affective support tailored to individual persona needs. Preliminary expert validation confirms pedagogical soundness and highlights key ethical considerations. The paper presents this integrative framework alongside a detailed discussion of its design principles, operative AI cycle, and embedded ethical safeguards, positioning persona-aware AI as a transformative, yet responsible, pathway for computational mathematics education.

Keywords. Computational Thinking, Generative AI, Learner Personas, Digital Mathematics Learning Environments (DMLEs), Scaffolding, Personalized Learning

1. Introduction

The imperative to embed Computational Thinking (CT) within mathematics education is clear, as it empowers students to become active modelers and analysts of dynamic systems [1, 2]. Despite this potential, the persistent reality in Digital Mathematics Learning Environments (DMLEs) is one of student frustration, disengagement, and cognitive overload [3]. The problem lies not in technology availability, but in the lack of adaptive, diagnostic, and supportive intelligence within these environments. Existing

[1] Corresponding Author, Allan M Canonigo, University of the Philippines, Philippines; Email: amcanonigo@up.edu.ph

DMLEs are often architecturally static, providing uniform feedback and paths that treat the individualized learning process as monolithic [4], thereby failing to address diverse cognitive and affective needs of learners.

Prior research has established that students approach computational problem-solving through distinct, recognizable learner personas characterized by specific epistemic stances and affective responses [4, 5]. However, the translation of these diagnostic profiles into scalable, real-time pedagogical interventions remains underexplored. Emerging Generative AI (GenAI) capabilities present a paradigm-shifting opportunity to bridge this gap. GenAI can generate dynamic, context-aware, and multi-modal responses, engineered to act as personalized "co-thinkers" in the learning environment [6]. This aligns with a broader conceptualization of AI as a collaborative educational agent [7].

This paper directly addresses the gap by moving from learner identification to principled design of transformative interventions. It poses the question: How can Generative AI be strategically engineered to create persona-aware scaffolds that transform debugging in DMLEs from a point of frustration into a personalized dialogue for mathematical sense-making? The primary contribution is a theoretically grounded framework integrating validated learner profiles with targeted GenAI prompt strategies. An initial expert-validated conceptual prototype demonstrates the framework's potential and outlines a clear path for future empirical testing. By theorizing AI as an adaptive metacognitive and affective partner, this research charts a course toward a new generation of DMLEs that are not just computationally powerful, but pedagogically intelligent and empathetically responsive [8].

2. Theoretical Framework and Literature Review

This investigation synthesizes established learning theories with the affordances of Generative AI (GenAI) to contribute to educational AI research. It constructs an integrated model that transcends isolated interventions by orchestrating multiple AI-supported strategies to address the cognitive, metacognitive, and affective dimensions of learning in computationally complex domains such as mathematics.

2.1 Constructionism and the Object-to-Think-With Paradigm

Constructionism posits that learning deepens most effectively through the active construction of public, shareable artifacts [9]. In Digital Mathematics Learning Environments (DMLEs), the algorithm or computational model traditionally serves as this core artifact. GenAI transforms this static object into a dynamic, interactive partner. Students can now engage in a rich, iterative "conversation" with the artifact—querying its logic, proposing modifications, and interpreting its behaviors in natural language [10]. This amplifies the object-to-think-with paradigm, as the AI-mediated artifact becomes a responsive entity that challenges assumptions, explains its own structure, and co-

constructs knowledge with the learner, thereby making abstract computational concepts more tangible and malleable.

2.2 Scaffolding Computational Thinking Practices

The core practices of Computational Thinking (CT)—such as debugging, abstraction, decomposition, and algorithm design—require the internalization of expert-like problem-solving heuristics. GenAI can scaffold these practices by explicitly modeling expert chains of thought, effectively externalizing the reasoning process that experts often tacitly follow [11]. By verbalizing this reasoning, GenAI makes invisible disciplinary practices explicit, apprenticing students into essential cognitive habits. This approach is consonant with the Productive Failure framework, which underscores the value of students grappling with complex problems before receiving direct instruction [12]. Within this struggle, GenAI can function as a sensitive detector of unproductive frustration, intervening with timely, targeted hints or questions. This capability helps maintain learners within their productive zone of proximal development, ensuring that challenge is balanced with actionable support to prevent disengagement.

2.3 GenAI as a Dual Metacognitive and Affective Partner

The framework conceptualizes GenAI as a dual-purpose pedagogical partner. Metacognitively, it fosters self-regulation by prompting learners to articulate reasoning and reflect on strategies [13]. Affectively, it responds to states like frustration or anxiety with empathetic encouragement and reframing [14]. While components like persona-aware tutoring [5], metacognitive prompting [13], affective feedback [14], and RAG for accuracy [15] exist in isolation, this work's novelty is their "integrated orchestration" for computational mathematics. It explicitly links learner profiles to specific prompt designs, creating a holistic and implementable scaffolding model.

3. Methodology

Employing a two-phase Design-Based Research (DBR) methodology focused on developing practical design knowledge and theoretical understanding through iterative design [16], this study aimed to validate a learner persona typology and subsequently translate it into a conceptual GenAI scaffold framework. A clear distinction is maintained between the empirical validation phase (Phase 1) and the conceptual design and expert validation phase (Phase 2).

3.1 Phase 1: Empirical Persona Validation

This phase constituted a qualitative study with the purpose of validating a typology of learner personas based on observable epistemic stance, affective response, and debugging behavior within a DMLE. The study was conducted with 28 secondary school

students engaged in a 3-week algebra unit focused on modeling linear functions using the Snap! block-based programming environment. Data were triangulated from four sources: (1) think-aloud protocols during structured debugging tasks, (2) semi-structured, in-depth interviews employing stimulated recall techniques (average length: 25 minutes), (3) analysis of iterative code artifacts, and (4) DMLE log data capturing behavioral metrics.

A hybrid inductive-deductive qualitative coding approach was applied to transcripts. The *deductive* codes were derived from the literature on epistemic stance (e.g., procedural, model-based) and affective pathways (e.g., anxiety → avoidance). *Inductive* codes emerged from the data, characterizing specific debugging behaviors (e.g., "monolithic testing," "random value changes"). The final persona categorization was based on the consistent clustering of three dimensions: the dominant epistemic stance, the primary affective pathway, and the signature debugging behavior. To support validity and reliability, coding and clustering were conducted independently by two researchers, achieving an inter-coder agreement of Cohen's kappa = 0.82. Discrepancies were resolved through discussion.

3.2 Phase 2: Development and Expert Validation of Conceptual GenAI Scaffolds

The purpose of this phase was to operationalize the empirical findings from Phase 1 into a structured, evidence-based design framework for AI-powered interventions. This was conducted as a formal design research process.

3.2.1 Empirical Derivation of Design Principles and Persona-Specific Scaffolding Goals

The analysis of Phase 1 data was systematically translated into a set of provisional design principles. For each of the four empirically derived learner personas, a core "Point of Struggle" was defined. Correspondingly, a targeted "AI Scaffolding Goal" was established. For instance, data showing that "Procedural Procrastinators" (later refined to "Conceptually Stalled Learners") consistently avoided algorithmic planning due to conceptual overwhelm led to the scaffolding goal: "To bridge abstract concepts to concrete code and re-engage with the task."

3.2.2 Development of Structured Prompt Templates as Conceptual Prototypes

Guided by these principles, we authored a suite of structured prompt templates as detailed conceptual prototypes. Each template defined precise conditional logic, a pedagogical function, and exact phrasing. For instance, the template for the Trial-and-Error Tinkerer is designed to activate after observing excessive, unfocused execution cycles (IF (execution_cycles > 30) AND (changes_are_numerical_only == TRUE)…). Its goal is to "channel unstrategic action into hypothesis-driven testing" using prompts such as: *"I notice you're actively exploring... To debug more efficiently, let's form one hypothesis first."* A fallback strategy is included for continued unfocused activity. Each

template was mapped to its corresponding persona, struggle point, and scaffolding goal, forming a testable "promptbook" artifact.

3.2.3 Expert Review Panel for Validation

To validate the framework and prototypes, an expert panel of three learning scientists and two AI ethics scholars assessed them using a structured protocol focusing on pedagogical soundness, technical coherence, and ethical risk. Feedback from questionnaires and a focus group validated the theory-practice linkage and identified key risks—such as reduced cognitive demand and privacy concerns—guiding prototype revision. This process ensured the designs were vetted for rigor and responsibility [6, 15] prior to technical implementation.

4. Findings and Framework Design

4.1 Validating the Four Learner Personas

The empirical analysis robustly confirmed a four-persona typology. Quantitative behavioral data from DMLE logs provided clear signatures that corroborated the qualitative findings. The persona definitions, their behavioral signatures (mean code execution cycles per task session), and the derived AI scaffolding goals are summarized in Table 1.

Table 1: Validated Learner Personas, Behavioral Signatures, and Corresponding GenAI Scaffolding Strategies

Persona	Epistemic Stance & Affective Pathway	Behavioral Signature (Mean Execution Cycles ± SD)	GenAI Scaffolding Goal	Prototyped GenAI Intervention
Frustrated Perfectionist	Stance: Procedural, seeks single correct answer. Pathway: Anxiety → Avoidance.	3.1 ± 1.4 (low, monolithic testing)	Affective Reassurance & Process Valorization	"I like how you've structured the main sequence. The loop error is a common chance to learn. Let's focus just on the loop's end condition."
Trial-and-Error Tinkerer	Stance: Experimental, surface-level. Pathway: Curiosity → Unstrategic Action.	48.7 ± 12.3 (high, random execution)	Metacognitive Prompting for Strategic Planning	"I see you're actively exploring! To solve this efficiently, what's your one hypothesis? Now, use 'Isolate & Test' on the slope block."
Systematic Debugger	Stance: Model-based, reflective. Pathway: Determination → Productive Struggle.	15.8 ± 4.7 (methodical, decomposed testing)	Semantic Trace Visualization & Conceptual Inquiry	"Step 3: x=2. Math Meaning: independent variable. Step 4: y=5. Output matches. Query: What if you changed one block to model exponential growth?"

Persona	Epistemic Stance & Affective Pathway	Behavioral Signature (Mean Execution Cycles ± SD)	GenAI Scaffolding Goal	Prototyped GenAI Intervention
Conceptually Stalled Learner	Stance: Disconnected, abstract. Pathway: Confusion → Overwhelm.	4.5 ± 2.1 (very low engagement, conceptual errors)	Conceptual Bridging & Re-connection	"Let's connect the math to the code. In y=2x+5, the 'y-intercept' (5) is the starting value. Which block sets the starting position?"

4.2 A Framework for GenAI Integration: The Detect-Select-Execute-Refine Cycle

The validated personas form the cornerstone of the proposed framework. The transition from persona diagnosis to intervention is conceptualized within a continuous Detect-Select-Execute-Refine cycle, transforming the DMLE into an interactive co-thinker.

- Detect: The system analyzes real-time learner activity (code edits, execution patterns, potentially natural language queries) and log data to infer the most probable current learner persona and cognitive-affective state.
- Select: Based on the detection output, the system selects the most appropriate pre-engineered prompt template from its library (as exemplified in Table 1 and Section 3.2.2).
- Execute: The GenAI model, potentially augmented with a Retrieval-Augmented Generation (RAG) system anchored to the verified curriculum, generates a final, context-specific response from the selected template.
- Refine: The system monitors the learner's subsequent interaction following the intervention to assess its effectiveness (e.g., did debugging become more focused? Did engagement increase?), using this feedback to refine future detection models and prompt selections.

This cycle is visualized in Figure 1. Persona-specific prompt templates form the core mechanism that links detection to pedagogical action.

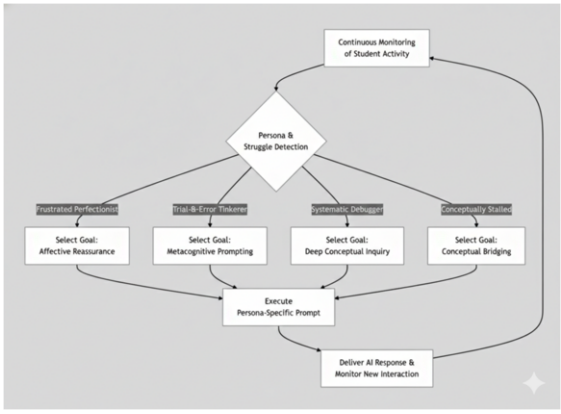


Figure 1: Framework for Persona-Aware GenAI Integration in DMLEs
(Note: This is an AI-generated diagram conceptualizing the Detect-Select-Execute-Refine cycle, with "Persona-Specific Prompt Templates" at the core linking learner activity to tailored AI responses.)

4.3 Informing the Framework with Real-World Agent Paradigms

The proposed framework draws practical validation from the proliferation of adaptive, conversational AI agents across industries, which demonstrate the core functionalities required for persona-aware scaffolding [6, 17]. For instance:

- Natural Dialogue Systems like automotive co-pilots [18] model the context-aware, engaging interaction essential for sustaining dialogue with diverse learners.
- Knowledge-Sector Copilots in fields like finance [19] exemplify AI that supports the decomposition of complex, knowledge-intensive tasks—directly mirroring the support needed for mathematical problem-solving.
- Retrieval-Augmented Generation (RAG) in customer support [20] provides a blueprint for delivering accurate, domain-grounded explanations, a technique directly applicable to mitigating AI hallucination in education by tethering responses to curricular content.

Collectively, these paradigms affirm a shift toward specialized AI agents that democratize access to expertise [19, 20]—a capability critical for equitable computational mathematics education.

5. Discussion

This research demonstrates that a synergistic approach, combining a nuanced understanding of learner diversity with the advanced capabilities of Generative AI, can address longstanding scalability barriers in personalized education. The primary contribution is a novel, integrative framework that provides a concrete design pathway for evolving static DMLEs into adaptive, persona-aware co-thinkers. It advances the field by offering a systematic model for translating learner psychology into actionable AI prompt architectures, thereby operationalizing calls for affective computing [14], metacognitive scaffolding [13], and persona-aware design [5] within a unified, domain-specific application.

The proposed interventions represent an evolution beyond traditional intelligent tutoring systems through their targeted, theory-informed design. The framework proactively engages with a central tension in human-AI collaboration: balancing helpful augmentation with the risks of cognitive overload or over-reliance [7]. By making the AI's support contingent on the learner's demonstrated persona and state, the design aims to scaffold the gradual development of independent problem-solving skills, adapting support dynamically as the learner evolves.

5.1 Ethical and Technical Imperatives Embedded in Design

A responsible implementation of this framework requires ethical safeguards to be structurally embedded, aligning with international guidelines such as the UNESCO

Recommendation on the Ethics of AI [8]. This entails mitigating hallucination and ensuring accuracy through a Retrieval-Augmented Generation (RAG) approach, anchoring responses to a verified curriculum database [15, 20], and designing prompts that model epistemic humility. To prevent bias, persona detection must be treated as a dynamic, probabilistic estimate—not a fixed label—with algorithms regularly audited for fairness. Finally, privacy and agency must be upheld through minimal, transparent data collection and by ensuring learners retain control over interactions, thereby designing for empowerment rather than surveillance.

6. Conclusion, Limitations, and Future Directions

This study has advanced a vision for computational mathematics education through a framework that maps empirically validated learner personas onto the responsive capabilities of Generative AI. The principal contributions are twofold: (1) the validation of a four-persona typology, supported by quantitative behavioral signatures, and (2) the development and expert validation of a design framework that translates these personas into structured GenAI prompt architectures.

Several limitations warrant acknowledgment. The proposed AI interventions remain conceptual prototypes; their efficacy and impact on learning outcomes necessitate empirical evaluation in authentic educational settings. Furthermore, the generalizability of the persona typology beyond the specific context of secondary algebra and block-based programming requires further investigation. Finally, practical implementation faces significant technical challenges, including the development of robust real-time detection algorithms and seamless Retrieval-Augmented Generation (RAG) systems.

These findings delineate clear trajectories for future work. Researchers are directed toward implementation and controlled mixed-methods studies to measure effects on computational thinking acquisition, mindset, and engagement. For instructional designers, the persona-scaffold mapping offers a practical blueprint for authoring adaptive content. At the policy level, this work underscores the imperative of teacher professional development for AI-augmented classrooms [21] and the establishment of robust ethical guidelines for educational AI [8]. Ultimately, the responsible evolution of such tools is contingent upon a committed, iterative co-design process that actively involves educators and learners at every stage [22].

References

[1] Wing JM. Computational thinking. Commun ACM. 2006;49(3):33–35. https://doi.org/10.1145/1118178.1118215

[2] Weintrop D, Beheshti E, Horn M, Orton K, Jona K, Trouille L, et al. Defining Computational Thinking for Mathematics and Science Classrooms. J Sci Educ Technol. 2016; 25(1):127–147. https://doi.org/10.1007/s10956-015-9581-5

[3] Papert S. Mindstorms: Children, Computers, and Powerful Ideas. New York: Basic Books; 1980.

[4] Isotani S, Inaba A, Ikeda M. User Modeling in Education: Concepts and Algorithms. User Model User-Adap Inter. 2019; 29:497–577. https://doi.org/10.1007/s11257-019-09238-5

[5] Govaerts S, Santos OC, Verbert K. A systematic review on learner and user modeling in education and training. User Model User-Adap Inter. 2019; 29:497–577. https://doi.org/10.1007/s11257-019-09221-1
[6] Zawacki-Richter O, Marín VI, Bond M, Gouverneur F. Systematic review of research on AI in education. Int J Educ Technol High Educ. 2019;16(1):1–27. https://doi.org/10.1186/s41239-019-0171-0
[7] Forlizzi J, Zimmerman J. Promoting innovation through design research. Des Issues. 2013; 29(3):5-14. https://doi.org/10.1162/DESI_a_00192
[8] UNESCO. Recommendation on the Ethics of Artificial Intelligence. Paris: UNESCO; 2023 [cited 2024 Mar 15]. Available from: https://unesdoc.unesco.org/ark:/48223/pf0000380455
[9] Luckin R. Machine Learning and Human Intelligence: The future of education for the 21st century. London: UCL Institute of Education Press; 2018.
[10] Papert S. Situating Constructionism. In: Harel I, Papert S, editors. Constructionism. New York: Routledge; 1991. p. 1-11. https://doi.org/10.4324/9781315084820
[11]Grover S, Pea R. Computational thinking in K-12. Educ Res. 2013;42(1):38-43. https://doi.org/10.3102/0013189X12463051
[12]Kapur M. Productive Failure. Cogn Instr. 2008; 26(3):379–424. https://doi.org/10.1080/07370000802212669
[13] D'Mello SK, Graesser A. Feeling, thinking, and computing with affect-aware learning technologies. In: Sottilare RA, Graesser AC, Hu X, Holden HK, editors. Design Recommendations for Intelligent Tutoring Systems: Volume 3. Orlando: U.S. Army Research Laboratory; 2015. p. 419-441. https://doi.org/10.1007/978-3-319-05318-4_22
[14] Picard RW. Affective Computing. Cambridge: MIT Press; 1997.
[15] Lewis P, et al. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. Adv Neural Inf Process Syst. 2020; 33:9459–9474.
[16] Brown AL. Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. J Learn Sci. 1992;2(2):141-178. https://doi.org/10.1207/s15327809jls0202_2
[17] Xu W, Dainoff MJ, Ge L, Gao Z. Transitioning to human interaction with AI systems: New challenges and opportunities for HCI professionals to enable human-centered AI. Int J Hum-Comput Interact. 2023;39(3):494-518. https://doi.org/10.1080/10447318.2022.2041900
[18] Yilmaz M, et al. AI-Driven Vehicle Control. IEEE Trans Intell Transp Syst. 2019;20(12):4505-4517. https://doi.org/10.1109/TITS.2019.2927154
[19] Agarwal B, Gautham L. AI in finance: applications and research directions. J Finance Data Sci. 2021;7(2):167-181. https://doi.org/10.1016/j.jfds.2020.12.005
[20] Searle RH, et al. AI chatbots in customer service: New barriers and opportunities. Int J Hum-Comput Stud. 2021; 146:102557. https://doi.org/10.1016/j.ijhcs.2021.102557
[21] Redecker C. European Framework for the Digital Competence of Educators: DigCompEdu. Luxembourg: Publications Office of the European Union; 2017. https://doi.org/10.2760/159770
[22] Eubanks V. Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. New York: St. Martin's Press; 2018.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261094

The Smart School Model: Integrating Intelligent Agents and IoT for Autonomous, Ethical and Reproducible Educational Management

Youssef Lahbabi*, AIT OUFKIR Abdellah, Lalla Touhfa BELGNAOUI and Tarik LAFOU
EMAA Business School-Agadir Morocco

Abstract. Smart-school projects increasingly combine sensors, dashboards and artificial intelligence, but their managerial value remains weak when attendance, access control, energy use and reporting operate as isolated modules. This paper proposes a Smart School Management System (SSMS) in which Internet-of-Things observations are transformed into auditable decisions through a policy-constrained multi-agent architecture. Unlike component-level smart-campus prototypes, the contribution is an orchestration model linking perception agents, decision agents and a coordination agent under privacy, traceability and human-oversight constraints. The empirical part uses the open UCI Occupancy Detection dataset (20,560 labelled records) to validate the perception layer on a reproducible occupancy-classification task. Trained on 8,143 records and tested on 12,417 records, logistic regression achieved 98.34% accuracy and 0.9659 F1-score, outperforming a simple light-threshold rule (94.36% accuracy, 0.8961 F1). These results validate only the occupancy-perception component under the stated open-data protocol; they do not validate the complete smart-school system or institutional deployment. The paper therefore advances a reproducible, ethically governed and practically deployable direction for autonomous educational management.

Keywords. smart school, Internet of Things, multi-agent systems, educational management, occupancy detection, responsible AI, reproducibility

1. Introduction

Smart-school projects increasingly combine sensors, dashboards and artificial intelligence, but their managerial value remains weak when attendance, access control, energy use and reporting operate as isolated modules. This paper proposes a Smart School Management System (SSMS) in which Internet-of-Things (IoT) observations are transformed into auditable decisions through a policy-constrained multi-agent architecture. The purpose is not to replace school professionals, but to support them with traceable coordination

*Corresponding author: Youssef Lahbabi, EMAA Business School-Agadir Morocco. E-mail: y.lahbabi@emaa.ma

mechanisms that perceive context, reason over operational objectives and recommend or trigger bounded actions.

The research problem is the lack of a reproducible, ethically governed architecture that links IoT sensing to adaptive coordination for daily school management. Existing smart-campus systems often emphasize connectivity and dashboards, whereas AI-in-education studies optimize single analytic tasks such as attendance, prediction or engagement detection [1,2]. Four gaps motivate this study: fragmented access, attendance, energy and reporting subsystems; brittle rule-based automation; weak reproducibility due to private data and vague baselines; and insufficient privacy, proportionality and human-supervision controls in sensitive school contexts. Accordingly, the guiding question is: how can a smart school architecture integrate open-data-driven IoT analytics and intelligent agents to improve operational coordination while preserving reproducibility, accountability and ethical safeguards?

This study deliberately separates architectural design from empirical validation. The SSMS is a governance-oriented reference architecture, whereas the executed experiment validates only its occupancy-perception component.

The paper contributes: (1) a four-layer SSMS architecture; (2) a policy-constrained multi-agent decision model; (3) a reproducible experiment on the UCI Occupancy Detection dataset; (4) a comparison between transparent rule-based and supervised baselines; and (5) privacy, traceability and human override as explicit design requirements.

2. Critical Related Work

IoT and smart-campus infrastructures provide sensing and connectivity, but often remain dashboard-centric: data are collected and visualized while cross-domain decisions depend on manual interpretation [3]. AI in education shows strong potential for monitoring, prediction, personalization and feedback, yet many studies remain learner-facing or single-task, which makes institutional deployment difficult [1,2]. Multi-agent systems are relevant because they distribute perception, decision and coordination across specialized entities and negotiate competing objectives such as energy efficiency, comfort, safety and oversight [4,5]. The originality of this paper lies in transferring that coordination logic to smart-school management while grounding a core perception module in open, reproducible data.

The resulting positioning is clear: the proposed SSMS does not add another isolated dashboard; it binds sensing, reasoning, action logging and governance into one auditable workflow.

3. Proposed Smart School Management System

Figure 1 presents the proposed SSMS architecture. The perception layer captures events and environmental conditions; the communication and data layer secures and structures streams; the intelligent agent layer transforms observations into decisions; and the application and actuation layer exposes outputs to administrators, teachers, parents and physical actuators. This separation prevents direct sensor-to-action shortcuts and makes decisions traceable.

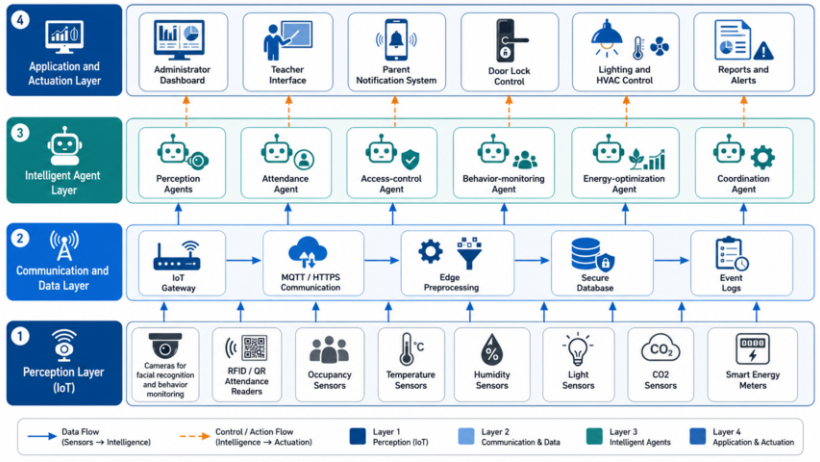


Figure 1. Layered architecture of the Smart School Management System.

The architecture avoids treating facial recognition or behavior monitoring as unconstrained surveillance. Biometric access, where legally permitted, should be optional, purpose-limited and subject to human review. Behavior-related signals are framed as risk or context indicators rather than disciplinary conclusions. The perception layer includes cameras, RFID or QR readers, occupancy sensors, indoor-air sensors and smart meters. The communication layer uses gateways, MQTT/HTTPS messaging, edge preprocessing, secure databases and event logs. The intelligent agent layer includes perception, attendance, access-control, behavior-monitoring, energy-optimization and coordination agents. The application layer provides dashboards, notifications, door control, lighting, HVAC control and reporting. A school can begin with non-biometric occupancy and energy sensors, then add other modules only after legal, ethical and organizational assessment.

4. Methodology and Open Data

The study follows a design-science methodology. It specifies the SSMS artifact, justifies it through critical comparison with prior work and defines a reproducible evaluation workflow. Figure 2 shows the workflow. Real open datasets are used to avoid unverifiable claims. Synthetic school-event logs are used only to emulate administrative events that public datasets rarely contain; they are not presented as real school data.

The executed numerical experiment uses the UCI Occupancy Detection dataset: 20,560 labelled indoor observations, with 8,143 training records, 2,665 observations in the first test file and 9,752 observations in the second test file [6]. The variables used are temperature, humidity, light, CO_2 and humidity ratio, with occupancy as target. Building Data Genome Project 2 and ASHRAE Thermal Comfort Database II support future energy-load and comfort scenarios [7,8]; they are not over-presented as executed experiments. Table 2 distinguishes the executed occupancy benchmark from datasets reserved for future energy-load and thermal-comfort scenarios, thereby making the empirical scope explicit.

Preprocessing and reproducible environment. The three UCI files were read with their published partitions. Automated checks found no missing values and no duplicate rows,

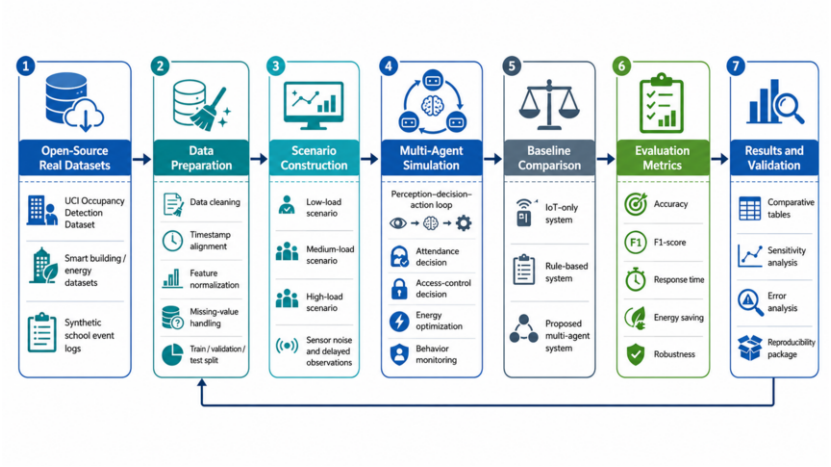


Figure 2. Research methodology and experimental workflow. Only the occupancy-perception benchmark is executed in the present study.

Table 1. Reproducible model and software settings.

Item	Setting
Data partition	Published training file ($n = 8,143$); two published test files concatenated only for external evaluation ($n = 12,417$)
Logistic regression	StandardScaler; `max_iter`=1000; default `lbfgs` solver; seed 42
Decision tree	`max_depth`=5; seed 42
Random forest	200 trees; seed 42; parallel execution
Gradient boosting	scikit-learn defaults; seed 42
Verification environment	Python 3.13.5; scikit-learn 1.8.0; NumPy 2.3.5; pandas 2.2.3; Linux

so no records were removed or imputed. The five predictors were retained in their original units. Logistic regression was fitted through a pipeline that standardizes predictors using statistics learned from the training partition only; tree-based models used the original scales. The class distribution was preserved without over- or undersampling. The verification run used Python 3.13.5, scikit-learn 1.8.0, NumPy 2.3.5 and pandas 2.2.3 on Linux, with random seed 42 for stochastic estimators. Table 1 reports the fixed settings.

5. Multi-Agent Decision Model

Each agent A_i is represented as

$$A_i = \langle S_i, O_i, P_i, D_i, \pi_i \rangle, \tag{1}$$

where S_i is the internal state, O_i is the observation set, P_i is the objective set, D_i is the decision space and π_i maps observations to actions. A decision is acceptable only if it satisfies the governance constraint set G, which includes safety, privacy, authorization and human-override rules:

Table 2. Open datasets and experimental use.

Dataset	Variables or role	Use in this paper
UCI Occupancy Detection	Temperature, humidity, light, CO_2, humidity ratio and occupancy	Executed room-state inference benchmark
Building Data Genome Project 2	Building-meter profiles and metadata	Future energy-load scenarios for lighting/HVAC policies
ASHRAE Thermal Comfort Database II	Comfort and indoor-condition descriptors	Comfort constraints for HVAC recommendations
Synthetic school event logs	Access requests, timetable slots and audit fields	Workflow coordination when no equivalent public school dataset exists

$$a_t^* = \arg\max_{a \in D_i} U(a \mid S_i, O_i, H_t), \quad \text{subject to } a \in G. \tag{2}$$

Thus, the highest-scoring action cannot be executed if it violates governance constraints. For example, an alert may be generated for review, but a punitive decision cannot be automated from a behavioral-risk signal.

Communication, conflict resolution and escalation. Agents exchange versioned JSON event objects over authenticated MQTT or HTTPS channels. Each object carries a unique event identifier, timestamp, source, confidence or quality indicators, proposed action and policy context. Decision agents publish recommendations rather than writing directly to actuators. The coordination agent groups recommendations by event identifier, rejects stale or unauthorized messages and applies the following precedence: (1) safety; (2) privacy and legal constraints; (3) human-approved school policy; and (4) operational objectives such as comfort and energy efficiency. Within the same priority level, the action with the highest policy-compliant utility is selected; unresolved ties or incompatible recommendations are escalated. Human approval is mandatory for access denial, disciplinary or behavior-related consequences, use of biometric evidence, policy exceptions, low-confidence observations and any case for which no action satisfies all constraints. Only predefined, reversible and low-risk infrastructure actions may be executed automatically, and every proposal, rejection, approval and execution is written to the audit log.

Illustrative occupancy-based HVAC decision. Suppose that an occupancy event indicates that a room has remained empty for 25 minutes. The perception agent first checks signal completeness and confidence. The energy agent proposes HVAC standby, while a timetable agent may report a scheduled examination and a comfort agent may request pre-conditioning. The coordination agent gives safety and approved timetable policy precedence over energy savings. It executes standby only when the room is confidently empty, no protected activity is scheduled and the action is reversible; otherwise it preserves the current state or routes the recommendation to the facilities manager. The final decision and its supporting observations are logged under the same event identifier. Figure 3 summarizes, in the main decision-model discussion, this governed perception-decision-action cycle and its human-approval gate.

At each time step, perception agents validate data quality and extract context; decision agents propose actions; the coordination agent checks conflicts and governance constraints; approved actions are logged and executed or routed to a human reviewer. Each incoming observation is assigned an event identifier and processed through validation, context

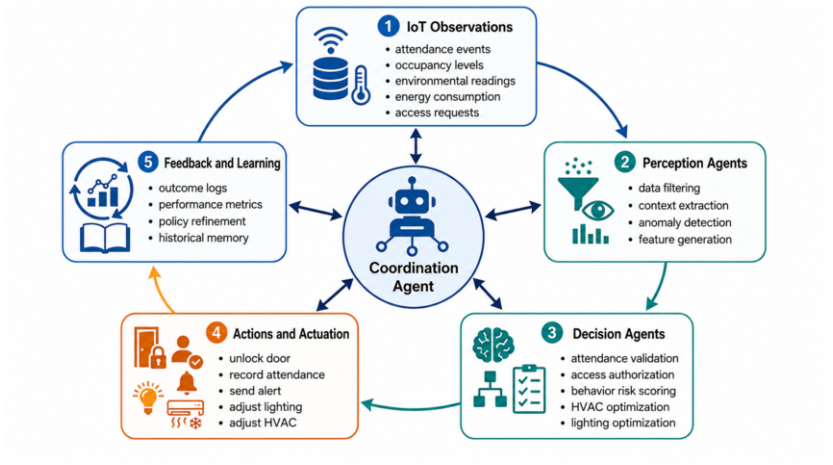


Figure 3. Multi-agent perception-decision-action feedback loop. Human approval is required whenever policy constraints, confidence requirements or action sensitivity prevent autonomous execution.

extraction, candidate-action generation, governance checking and logging. This allows the school to audit which sensor values, policy rules and human decisions produced a given outcome.

6. Empirical Evaluation on Open Data

The training split was `datatraining.txt` ($n = 8,143$), while `datatest.txt` and `datatest2.txt` were concatenated as an external test set ($n = 12,417$). The baseline was a transparent rule, Light > 100. Four supervised models were trained with fixed random seeds and no manual tuning beyond standard scaling for logistic regression. The purpose is not to optimize a leaderboard score; it is to verify whether a core perception agent can produce reproducible evidence before school-level actuation.

Robustness was evaluated without retraining. First, the fitted logistic model was scored separately on each published test file. Second, independent zero-mean Gaussian perturbations were added to every test predictor at 1%, 5% and 10% of that predictor's training-set standard deviation, using NumPy generator seed 42; metrics were then recomputed on the perturbed test sets. To examine dependence on light, the same standardized logistic-regression specification was also fitted with Light alone and with all predictors except Light.

Table 3 compares the transparent threshold baseline with four supervised models on the unchanged external test set.

Logistic regression achieved 98.34% accuracy and 0.9659 F1-score on the combined external test set. Its confusion matrix was TN = 9,290, FP = 106, FN = 100 and TP = 2,921. This matters because false occupancy decisions affect attendance logs and energy actions. The rule baseline achieved high recall but produced 697 false positives, showing why simple thresholds may keep empty rooms classified as occupied. The random forest was competitive but slower and less transparent. The evidence therefore supports interpretable supervised perception models for low-risk room-state decisions, while keeping sensitive actions subject to coordination-agent checks and human review.

Table 3. Empirical occupancy-classification results on the UCI dataset.

Model	Accuracy	Precision	Recall	F1-score
Rule-based (Light>100)	0.9436	0.8124	0.9990	0.8961
Logistic Regression	**0.9834**	**0.9650**	0.9669	**0.9659**
Decision Tree	0.9029	0.8646	0.7123	0.7811
Random Forest	0.9678	0.9143	0.9573	0.9353
Gradient Boosting	0.9482	0.9483	0.8325	0.8867

Table 4. Light-dependence analysis using logistic regression.

Predictor set	Accuracy	Precision	Recall	F1-score
All five predictors	0.9834	0.9650	0.9669	0.9659
Light only	**0.9893**	0.9619	**0.9954**	**0.9784**
All predictors except Light	0.7882	0.5669	0.5482	0.5574

Table 5. Robustness and action-level validity checks.

Check	Empirical result	Implication for SSMS
External split stability	`datatest` F1=0.9694; `datatest2` F1=0.9643	Not fitted only to one test file
Noise robustness	10% sensor-SD stress: accuracy=0.9795, F1=0.9576	Stable enough for advisory use
Empty-room action validity	Empty-action precision=98.94%; missed-saving rate=1.13%	Supports low-risk lighting/HVAC recommendations
Governance boundary	False-empty rate among occupied cases=3.31%	Sensitive actions must remain overrideable

Table 4 shows that performance depends strongly on Light in this office-room dataset: a light-only logistic model slightly exceeds the five-feature model, whereas removing Light causes a large decline. The multi-sensor result must therefore not be interpreted as evidence that all five variables contribute equally. Light is a plausible occupancy proxy in the source rooms, but its relationship to occupancy may change in classrooms with daylight, projection, cleaning, after-hours activities or different lighting policies. This dependency reinforces the need for school-specific validation and sensor-drift monitoring before operational use.

The robustness findings are summarized in Table 5, which reports split-specific stability, controlled-noise stress, action-level validity and the governance boundary associated with false-empty predictions.

7. Discussion, Governance and Limits

The UCI experiment validates room-state perception, split generalization and robustness under controlled noise; it does not validate facial access control, classroom behavior interpretation or institutional adoption. Occupancy data collected in office-like rooms may not reflect classroom movement, timetable structure, daylight exposure, lighting policy, cleaning periods or age-related behavior. The light-dependence analysis further shows

that the reported performance is partly dataset-specific. The framework should therefore be read as a rigorous, empirically grounded pre-deployment model rather than as a claim of universal school performance.

Governance is a design requirement, not an appendix. Schools process sensitive data, often involving minors. The system must apply privacy by design: purpose limitation, data minimization, encryption, role-based access, short retention, audit logs and human appeal. Biometric or behavior-related modules require prior legality, necessity, proportionality and bias assessment. The European AI Act, GDPR, NIST AI RMF and education-specific AI guidance provide guardrails for this logic [9,10,11,12]. Accountable-AI scholarship further supports model documentation, dataset documentation, explainability interfaces and internal audit procedures [13,14,15,16].

Deployment should be staged. Phase 1 uses non-biometric sensors and energy meters to test data quality and staff acceptance. Phase 2 adds advisory agents for lighting, HVAC and room-use reporting. Phase 3 enables controlled low-risk infrastructure actuation. Sensitive access or behavior-related modules should be considered only after privacy impact assessment, bias checks, lawful basis and appeal procedures.

8. Conclusion

This paper proposed an SSMS that integrates IoT sensing with intelligent agents for autonomous yet supervised educational management. Its main contribution is the shift from fragmented smart-school components toward a coordinated, traceable and policy-constrained architecture. On 20,560 open UCI occupancy records, a reproducible benchmark showed that logistic regression reaches 98.34% accuracy and 0.9659 F1-score on an external test set, outperforming a simple threshold rule. The result validates only the occupancy-perception component under this office-room dataset and experimental protocol, not the complete SSMS or its institutional effects. The model should be adopted progressively, beginning with non-biometric sensing, advisory recommendations and human oversight before sensitive automation. Future work will extend the benchmark to real school pilots, network-delay stress tests, energy-saving estimation and fairness audits for context-dependent decision policies.

Data Availability, Ethics and Reproducibility Statement

This manuscript uses only open, non-identifiable datasets. The executed experiment used UCI files `datatraining.txt`, `datatest.txt` and `datatest2.txt`. The accompanying script was rerun from the supplied UCI files and reproduced the reported model metrics, split-specific checks, noise-stress results, action-level risk indicators and confusion matrices; it also exports the Light-only and no-Light analyses. No real student records, facial images or biometric data are processed in the evaluation. Future deployment must obtain institutional approval, document lawful basis and retention periods, run bias and privacy-impact assessments and keep human decision-makers accountable for sensitive actions.

References

[1] Ouyang F, Jiao P. Artificial intelligence in education: The three paradigms. Comput Educ Artif Intell. 2021;2:100020. doi:10.1016/j.caeai.2021.100020.

[2] Chen L, Chen P, Lin Z. Artificial intelligence in education: A review. IEEE Access. 2020;8:75264-75278. doi:10.1109/ACCESS.2020.2988510.

[3] Atzori L, Iera A, Morabito G. The Internet of Things: A survey. Comput Netw. 2010;54(15):2787-2805. doi:10.1016/j.comnet.2010.05.010.

[4] Wooldridge M. An Introduction to MultiAgent Systems. 2nd ed. Chichester: Wiley; 2009.

[5] Russell S, Norvig P. Artificial Intelligence: A Modern Approach. 4th ed. Pearson; 2021.

[6] Candanedo LM, Feldheim V. Accurate occupancy detection of an office room from light, temperature, humidity and CO2 measurements using statistical learning models. Energy Build. 2016;112:28-39. doi:10.1016/j.enbuild.2015.11.071.

[7] Miller C, Kathirgamanathan A, Picchetti B, et al. The Building Data Genome Project 2, energy meter data from the ASHRAE Great Energy Predictor III competition. Sci Data. 2020;7:368. doi:10.1038/s41597-020-00712-x.

[8] Földváry Ličina V, Cheung T, Zhang H, et al. Development of the ASHRAE Global Thermal Comfort Database II. Sci Data. 2018;5:180105. doi:10.1038/sdata.2018.105.

[9] European Parliament and Council of the European Union. Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union. 2024.

[10] National Institute of Standards and Technology. Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST AI 100-1. 2023. doi:10.6028/NIST.AI.100-1.

[11] European Commission. Ethical guidelines on the use of artificial intelligence and data in teaching and learning for educators. Publications Office of the European Union; 2022.

[12] European Union. Regulation (EU) 2016/679, General Data Protection Regulation. Official Journal of the European Union. 2016.

[13] Mitchell M, Wu S, Zaldivar A, et al. Model cards for model reporting. In: Proceedings of the Conference on Fairness, Accountability, and Transparency; 2019. p. 220-229. doi:10.1145/3287560.3287596.

[14] Gebru T, Morgenstern J, Vecchione B, et al. Datasheets for datasets. Commun ACM. 2021;64(12):86-92. doi:10.1145/3458723.

[15] Amershi S, Weld D, Vorvoreanu M, et al. Guidelines for human-AI interaction. In: Proceedings of CHI; 2019. p. 1-13. doi:10.1145/3290605.3300233.

[16] Raji ID, Smart A, White RN, et al. Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In: Proceedings of FAT*; 2020. p. 33-44. doi:10.1145/3351095.3372873.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)

© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261095

Towards A Click'N'Tick Global Culture: Inevitable Future in Digital Age

Ljupcho EFREMOV[1]
Liberal Arts Department, American University of the Middle East, Egaila 54200, Kuwait
ORCiD ID: Ljupcho Efremov https://orcid.org/0000-0002-2660-4422

Abstract. The paper examines how global influences contribute to the formation of a global culture in the digital age. The study focuses on the similarities that have emerged as a result of globalization and aims to identify the elements which underlie a global culture. A scoping review combined with inductive discovery was used to identify cultural patterns in recent research. By analyzing existing studies, the paper sheds light on how globalization is shaping lifestyles. The review included 20 relevant studies from the last 5 years, with a focus on research directly related to elements of global culture. From this review, four intertwined elements of an emerging global culture were identified: Self-focused lifestyle (individualist and consumerist life); Digital communication and English as glue; Technology, AI and Fast Life and Common Artefacts and Infrastructures. The author named this type of Click'N'Tick, referring to the shift toward a digital world ("click") and a fast, time-pressured lifestyle ("tick"). Although these are four underlying elements, the studies also show that the global environment remains complex, with many layers of culture and important differences between groups and regions. Based on these findings, the paper proposes new label for the phenomenon and points the need to explore this trend in future research.

Keywords. Individualism; Self; English; Materialism, Consumerism, Technology

1. Introduction

Globalization is spreading more and more, especially in the digital age and the rise of AI. People around the world are using the same products and services, and technology and the internet have strongly accelerated these practices. The idea of a "global village" is now very visible in almost every sphere of life. People in the most remote areas have access to similar digital practices and use the same products. Apart from usage, digital communication prevails (even among people ling in the same neighborhood). A word that can describe this phenomenon is "Ubiquitous" which means something that seems to be everywhere at the same time or is very commonly found. Global companies and international organizations are creating and executing their strategies on global levels for many years. One study [1] points out that one of the clearest and most important advantages of globalization is easier access to international markets. Through the reduction of trade barriers and the fast growth of digital commerce platforms,

[1] Corresponding Author: Ljupcho Efremov, Liberal Arts Department, American University of the Middle East, 54200 Egaila, Kuwait; E-mail Ljupcho.efremov@aum.edu.kw

companies from large multinationals to very small firms can now work beyond their national borders. E-commerce platforms such as Amazon, Alibaba, Shopify, and Flipkart have changed how products are marketed and sold, because they offer real-time global reach with very little need for physical infrastructure. Another study [2] shows that digital marketing prevails over traditional marketing and the communication strategies of manufacturers are mostly online, through social media, online portals, and websites. The global culture trend started with "McDonaldization". That model [3] emphasized efficiency, predictability, calculability, and control, and through that it creates standardized experiences that put convenience before individuality. As McDonald's has expanded around the world, local food cultures are often pushed into the background by the attraction of fast food, which contributes to the loss of traditional culinary practices and tastes. This is just an example how global brands can reshape local cultures and preferences. However, the model which stared as unified experience and search for more efficiency is now spreading around the world as a lifestyle. The paper aims to look more deeply into this issue by exploring the elements that make up global culture.

2. Global Culture

The following part provides an overview of definition of global culture and explores different approaches and perspectives on globalization. First, globalization [4] means stronger interactions across borders and growing interdependence between countries in economic, cultural, technological, social, institutional, and ecological areas. The author writes that it started centuries ago with the European voyages of discovery, and it became much faster in the sixteenth and seventeenth centuries with European colonial expansion.

Related to globalization, a typical question about its influence is if a global culture exists. For this purpose, the paper starts with the perspective of Arnould and Thompson. Arnould and Thompson [5] created the Consumer Culture Theory (CCT). Instead of seeing culture as one big, uniform system that all members of a society share (for example, "Americans have this culture, Japanese have that culture"), CCT looks at how meanings are spread in a more uneven way and how many different, overlapping cultural groups exist within the larger context of globalization and market capitalism. Arnould and Thompson describe four main domains of CCT. First, they look at consumer identity projects, that is, how people use market offerings to build their identities and create shared worlds around common consumption interests. Second, they analyze marketplace cultures, meaning specific cultures that grow around brands, products, or leisure activities (for example, a sport like skydiving), where consumers act as culture producers. Third, they focus on the socio-historic patterning of consumption, looking at how wider institutional and social structures shape consumption, and how consumers' experiences, beliefs, and practices interact with these structures. Fourth, they examine mass-mediated marketplace ideologies and how consumers interpret, use, and sometimes resist these ideologies in their everyday lives. The critique to this perspective is about the growing tendency among CCT scholars [6] to see consumer culture not only through consumers' individual and collective agency, but also as something that is built into, and shaped by, the wider market system.

Related to the question if a global culture exists, the work of Goodman [7] is used as a main theoretical foundation for this paper. Goodman [7] argues that to understand global culture, authors need a definition of culture that recognizes how practices create meanings. Meaning-making practices also create the system that lies at the center of culture. The main focus should be on the practices that build this cultural system. At the end, Goodman concludes that global culture is therefore both a system and the set of practices that make up this system. Also, he talks about homogenization and heterogenization of global culture. Homogenization means the trend to sameness and less diversity in cultures around the world. This homogenization is sometimes called globalization, "Coca-colonization" or "McDonaldization". But not all these terms are equally good for describing this trend in global culture. Where there is homogenization of global culture, it is a special kind. It is not the spread of American culture or a soft drink or even a rationalizing process. It is the spread of consumer culture. Global forces don't make identical cultural things. Instead, they create a framework for new understanding of differences inside and between societies. Identity and difference are channeled into consumer practices. On the other hand, in contrast to global sameness, Mridha [8] writes about cultural heterogenization. It is not just that cultures stay diverse, but globalization increases diversity among cultures and especially inside them. On the other hand, Goodman claims that with globalization, Western cultural things are "indigenized" and get new local meanings; cultures influence each other and make new hybrids; Western culture itself gets more influenced by peripheral cultures; and new transnational cultural groups appear. One trend for more heterogeneity is called glocalization. Goodman says that the definition of culture helps to see that homogeneity and heterogeneity in consumer culture are not opposite. Instead, these common differences make the system of global culture. Consumer practices create and keep this system. Consumption gives chances for meaningful expression and resources for identity and social position. De Mooij [9] explains global culture as a process of developing a global understanding, in which the world becomes so interconnected that it is increasingly seen as one single place.

Another important perspective is the work of De Mooij and Hofstede [10]. De Mooij and Hofstede [10] discuss different areas of research using their Cross-Cultural Consumer Behavior Framework. In this framework, they organize the cultural aspects of the person into consumer attributes and processes, and the cultural aspects of behavior into consumer behavior domains. Income is treated separately: if there is no income, there is almost no consumption, so income is placed in its own box. The attributes of the person are about what people are (the who), while the processes are about what drives people (the how). The key question is "Who am I?", that is, how people describe themselves and others in terms of personality and identity; attitudes and lifestyle are closely connected to this. How people think, perceive, and what motivates them - how these aspects of "me" are turned into behavior. Their framework also includes implications for global branding and advertising.

Taking the abovementioned perspectives into account (especially Goodman's work), the cultural practices that constitute global culture need to be identified. To contribute to this discussion, this study tries to identify similarities among people related to global culture. Building on previous work and these arguments, the main aim of this paper is to examine the underlying elements of global culture in relation to lifestyle.

3. Methodology

The research technique employed in the paper is a scoping review, combined with inductive discovery. Scoping review [11] was chosen to align with the research question which is broad and the goal is to find and map all the available and emerging evidence on the topic of interest. A scoping review follows a series of structured steps: defining clear aims and research questions, planning a protocol, systematically searching the literature, screening and charting the studies, then analyzing and presenting the findings. Having clear aims and questions at the start is essential so that all these steps are coherent and focused. The goal of this study was to identify the common elements which underlie a global culture. The inclusion criteria focused specifically on studies written in English that researched global culture and provided information about elements of global culture. In this paper, culture is understood as a transmitted pattern of meanings embodied in symbolic practices, through which people communicate, maintain and transform their shared ways of life. Cultural practices are therefore taken as routines and activities (for example, media use, consumption, work, rituals) that produce and reproduce these meanings in everyday life. This definition is used when searching the literature and doing the framework development. The search terms were centered on "global culture" and "globalization" in relation to lifestyle. In contrast, the exclusion criteria removed studies about global culture that did not discuss specific cultural elements in the text and studies related to previous classifications. The review identified 27, which were later decreased to 20 papers, which are presented below, and most of them were found through Google Scholar. The analysis was focused on finding patterns that appeared in the material. The process relied on grounded theory. Grounded theory is defined by Charmaz [12] as "a method of conducting qualitative research that focuses on creating conceptual frameworks or theories through building inductive analysis from the data". The 'Click'N'Tick' model should therefore be interpreted as an exploratory and theoretical grounded framework.

4. Results - Review of elements of global culture in literature

The following part organizes the papers found through the scoping review. The studies that explore the relationship between globalization and elements of global culture are shown in a table which is organized around the four patterns (Papers are added and numbered in the table according to the order in which they were identified in the scoping review). Afterwards the main findings elaborating on the four elements are described.

Table 1. Categorization of studies according to the four patterns

Title	Key studies
Self-focused (Individualist and consumerist life)	Globalization's Impact On Commerce: Opportunities And Challenges [1]; Globalization And Cultural Homogenization: A Critical Examination [3]; Globalization and Cultural Homogenization: A Historical Perspective in Nigeria [17]; Globalization and Cultural Transformation: The Erosion of Traditional Values in Nigeria [18]; Globalization and Local Cultures: A Complex Coexistence [19];

	Globalization, Mobility, Identity, and Consumerism: an Analysis of the Genesis of Unsustainable Consumption [21]; Globalization Of Consumer Culture: An Empirical Survey Of Consumers In Delhi [22]; The Sharing of Culture: Global Consumerism [25]
Digital communication and English as glue	Cultural Diversity and Convergence in a Global Communication Era [15]; Bridging Cultures in Virtual Workplaces: A Communication-Focused Review of Global Virtual Teams [16]; The Impact of Globalization and Modern Technology on the English Language [26]; English Language and Globalization [27]; Hyper-centrality of English Language in the World of Globalization [28]
Technology, AI and fast life	The impact of technological advancement on culture and society [23]; AI Propulsion of Globalization [24]; 2025 World Investment Report [30]
Artefacts and common structures	Homogenization Or Diversification? The Impact Of Globalization On Cultural Identity Of The First And Second-Generation Immigrants [13]; Unity in diversity: navigating global connections through cultural exchange [14]; Cultural Identity and Globalization: The Impact on Indigenous Traditions [20]; The mobile phone as the globalizing icon of the early 21st century [29]

The analysis shows four distinctive patterns: Self-focused (Individualist and consumerist life), Digital communication and English as glue, Technology, AI and fast life and Artefacts and common structures. The text below explains the findings.

The Self-focused (individualist and consumerist life) element is about lifestyle and values. Globalization pushes lifestyles toward more standardized, often Western-oriented patterns, especially among youth, supporting individualism, materialism, and consumerism. These changes often come with economic growth and modern infrastructure but also a shift away from communal, spiritual, and tradition-based identities. Several studies show a move from community-oriented values toward more self-focused and consumption-driven lifestyles, especially in urban and globalized settings. The shift is particularly visible among younger generations, who increasingly embrace personal choice, career mobility, and brand-driven identities. The digital communication and English as a glue puts accents on English as lingua franca and online communication. Digital media, global communication networks, and virtual teamwork create shared spaces where similar practices, norms, and languages (often English) dominate. Online platforms and communication tools become key for cooperation and coordination across cultures, reinforcing a common "online" layer of everyday life and contributing to digital identities. Technology, AI and fast life is a self-explanatory term. Technological progress, media globalization, and new communication tools accelerate cultural change and normalize a fast, on-demand lifestyle. Emerging technologies, including AI and other digital innovations, support new forms of cultural exchange and convergence, but also intensify exposure to dominant global values. AI and fast life are grouped together as both are related to fast dynamics happening in the field of technology which is reflected in everyday life. The last pattern is about artefacts and common structures. Many studies highlight mobile phones as key symbols of globalization and central artefacts in global consumer culture. Mobile phones, global brands, and transnational corporate and supply-chain networks operate as shared material anchors of global culture, standardizing practices and expectations in daily life worldwide. The author named this phenomena as Click'N'Tick, referring to two main unifying shifts. The first one is about a digital world ("click") which includes digital communication, technology and artefacts. On the other hand, the second one puts the focus on efficiency related to self perspective in a

fast, time pressured lifestyle ("tick") unifying these patterns. The abovementioned term Click'N'Tick incorporates the four underlying elements of culture.

5. Conclusions

The results show the emergence of global culture named of Click'N'Tick. They are in line with Goodman's [7] view who argues that if a global culture does emerge, it will mainly take the form of a consumer culture, and this idea should guide empirical research to focus on concrete consumption practices that create and support such a culture. In the light of this discussion, the present study explains the main practices of global culture. The four main practices that emerged of the paper as elements of global culture are: Self-focused (Individualist and consumerist life), Digital communication and English as glue, Technology, AI and fast life and Artefacts and common structures. These practices outline common elements of global culture. The four proposed patterns reflect a new framework which is grounded in the four domains of CCT as proposed by Arnould and Thompson [5]. Namely the pattern of self focus revolves around identity and creation of shared consumption universe of CCT. The second domain of CCT which revolves around brands, products, or leisure activities is reflected in digital communication and emergence of English in a world where mass communication is the norm and acts as a producer of new culture. The third CCT domain (socio-historic patterns of consumption), looking at how wider institutional and social structures shape consumption is reflected in artefacts and common structures. The last but not least is about mass-mediated marketplace ideologies and how consumers interpret ideologies and this aspect is deeply rooted and shaped by technology and AI. The conclusion is that the four patterns (self-focused individualist and consumerist life, digital communication and English as glue, technology/AI/fast life, and artefacts/common structures) are reflected across their domains of consumer identity projects (self-focused practices), marketplace cultures (digital communication and tech as communal ties), socio-historic patterns of consumption (artefacts shaping structures), and mass-mediated marketplace ideologies (AI/fast life). However, the novelty of the proposed Click'N'Tick practices offers a more crystallized overview of these common global cultural elements. Overall the results are consistent with the findings of Enev et al. [31], who argue that, through mass media and the internet, Western brands and the English language are now widely used in advertising worldwide and have become symbols of modernity and internationalism for many consumers. Additionally, the proposed themes align with the work of Bartoli [32] about consumer self-concept and digitalization. The self-focused individualist and consumerist life practice aligns with consumer self-concepts centered on personalized digital narratives and identity-driven consumption in virtual spaces. Digital communication and English as glue mirrors the study's emphasis on constant connectivity. Technology, AI, and fast life correspond to human-technology assemblages, with algorithms shaping rapid, agentic experiences in identity formation. Artefacts and common structures parallel the use of digital/physical objects and brands. Another similarity is found in the studies of Makri et al. [33], who link global-local identities to digital consumption practices. Digital communication/tech is reflected in social networking sites as platforms for identity enactment and global affiliation. Self-focused consumerism is reflected through

possession of brands which signal global self via possessions/digital tools. Last, but not least, identities are link to digital consumption, reflected in AI/fast life and artefacts as practices of global culture. Another aligned study is found in work of Steenkamp [34] who arguments that global and local consumer cultures are grounded in consumer culture theory, cultural globalization theory and acculturation theory. The author claims that global consumer culture standardizes certain aspects of life, while people fuse global with local practices. A similar study by Ruževičiūtė and Ruževičius [35], argues that globalization reaches many domains of life and is reflected in buying similar products. However, the authors stress that even though people across the world buy the same products, their underlying motives are different (different national, religious or ethnic value systems).

This paper adds to the general knowledge on the topic by making a theoretical contribution to the field of "global culture" by introducing the term Click'N'Tick and identifying common practices. Although these are four underlying elements, the studies also show that the global environment remains complex, with many layers of culture and important differences between groups and regions. The limitations of the study are related to selected number of papers, while the conclusion lacks empirical validity. Alternatively, similar study can be conducted on more papers to check the conclusions. The empirical validity can be resolved through studies which explore the elements on a worldwide level. Also, the four elements of the model should be understood as an theoretically grounded and exploratory framework that would benefit from further testing and replication by independent research teams. One limitation of the study is that it is valid for middle and high class which adopts or strives towards adoption of lifestyle related to Click'N'Tick culture (similar to the concept of no income proposed by De Mooij and Hofstede [10]). Future studies should empirically test the proposed construct and evaluate discriminant validity. Comparative research designs would be useful to test the explanatory power of the proposed classification. Because the model is developed mainly from Western-centric perspectives and builds on Goodman's practices and operationalized definition, future work should test how well it generalizes to other cultural settings and explore possible contextual variations.

The study provides a unique term for global culture which relates to the four elements. The study incorporates AI as one of four main elements of culture as more and more people around the world use it. One study [36] shows that AI tools are used in everyday work tasks on a daily basis. Furthermore, besides theoretical, the study has practical implications. The study provides suggestions on how managers, politicians, and stakeholders can use cultural elements in their work. Managers of global companies, politicians and other relevant stakeholders (international organizations), can use the four elements to align offers and communication. The design of products and services can be tailored towards the proposed patterns while communication can use these elements to reach members of groups who value these features. However, additional research is needed to validate the four above-mentioned elements and to explore the direction and consequences of the trend and in which directions it will develop. Nevertheless, the study provides extraordinary background and sets a clear path for further researches related to global culture.

References

[1] Devi DJR. Globalization's impact on commerce: opportunities and challenges [Internet]. Int J Creat Res Thoughts. 2025 [cited 2026 Mar 19];13(6):1–6. Available from: https://www.ijcrt.org/papers/IJCRT25A6263.pdf

[2] Efremov L, Pruthi F. Evolution of Digital Marketing Among Automotive Sector in Kosovo. Economic Recovery, Consolidation, and Sustainable Growth [Internet]. 2023;619–35. Available from: http://dx.doi.org/10.1007/978-3-031-42511-0_40

[3] Amin T. Globalization and cultural homogenization: A critical examination. Kashf Journal of Multidisciplinary Research 2024; 1: 10–20.

[4] Okwir NA. Globalization and Its Impact on Local Cultures. Eurasian Experiment Journal of Arts and Management (EEJAM). 2025;7(2).

[5] Arnould EJ, Thompson C. Twenty years of consumer culture theory: Retrospect and prospect. Advances in Consumer Research. 2005;32:129.

[6] Askegaard S. Consumer culture theory (CCT). In: Cook DT, Ryan JM, editors. The Wiley Blackwell encyclopedia of consumption and consumer studies. Hoboken (NJ): Wiley-Blackwell; 2015. p. 1–4.

[7] Goodman DJ. Globalization and Consumer Culture. In: Ritzer G (ed) The Blackwell Companion to Globalization. Malden (MA): Blackwell; 2003. p. 330–51.

[8] Mridha MdAH. Global Culture: Its Existence and Consequences on Our Lifestyle in Bangladesh. Open Journal of Social Sciences [Internet]. 2021;09(01):167–87. Available from: http://dx.doi.org/10.4236/jss.2021.91012.

[9] de Mooij M. Consumer behavior and culture: consequences for global marketing and advertising. 3rd ed. London: SAGE Publications; 2019. 1–472.

[10] De Mooij M, Hofstede G. Cross-cultural consumer behavior: A review of research findings. Journal of international consumer marketing. 2011 May 1;23(3-4):181-92.

[11] Rodger D, Admani A, Thomas M. What is a scoping review? Evid Based Nurs 2024; 27: 84–85.

[12] Charmaz K. Constructing grounded theory: A practical guide through qualitative analysis. Sage; 2006 Jan 13.

[13] Bouymaj I, de Freitas Barbosa Pereira RT. Homogenization or diversification? The impact of globalization on cultural identity of the first- and second-generation immigrants. J Glob Stud. 2022;13(1):xx–xx. doi:10.30884/jogs/2022.01.06.

[14] Mariyono D, Alifatul Kamila AN, Alif Hidayatullah AN. Unity in diversity: navigating global connections through cultural exchange. Quality Education for All 2025; 2: 114–137.

[15] Guo L. Cultural Diversity and Convergence in a Global Communication Era. Epub ahead of print 8 April 2023. DOI: 10.5281/ZENODO.7810066.

[16] Liu X. Bridging Cultures in Virtual Workplaces: A Communication-Focused Review of Global Virtual Teams. Business and Professional Communication Quarterly 2025; 23294906251327747.

[17] Abdullahi I. Globalization and Cultural Homogenization: A Historical Perspective in Nigeria. IJHR 2024; 4: 51–63.

[18] Eyina NN. Globalisation and cultural transformation: the erosion of traditional values in Nigeria. In: Education in a Changing World: Humanities, Management, and Global Perspectives Conference; 2025 [Internet]. p. 48–58. Available from: https://top-conferences.us/index.php/ECHWMPGP/article/view/1994

[19] Wheatley Research Consultancy, Bagley, Minnesota, USA, Wheatley M. Globalization and Local Cultures: A Complex Coexistence. PJSS; 2024. Epub ahead of print 2024. DOI: 10.70389/PJSS.100005.

[20] Yılmaz N. Cultural identity and globalization: the impact on indigenous traditions. Arts Social Sci J. 2025;16(1):644. doi:10.37421/2151-6200.2025.16.644.

[21] Chintha SS, George B. Globalization, mobility, identity, and consumerism: an analysis of the genesis of unsustainable consumption. Palermo Bus Rev. 2012;(6):41–52.

[22] Bansal D. Globalization of consumer culture: An empirical survey of consumers in Delhi. Business Analyst. 2017;38(1):213-39.

[23] Alsaleh A. The impact of technological advancement on culture and society. Sci Rep. 2024;14(1):32140. doi:10.1038/s41598-024-83995-z.

[24] Drozd A. AI propulsion of globalization: the driving impact of artificial intelligence on global expansion [Internet]. Lynchburg (VA): Liberty University; 2024 [cited 2026 Mar 19]. Available from: https://digitalcommons.liberty.edu/cgi/viewcontent.cgi?article=2468&context=honors.

[25] Byrnes K. The sharing of culture: global consumerism. UW-L J Undergrad Res. 2007;10:1–10.

[26] Rahmani H, Karimi MI. The Impact of Globalization and Modern Technology on the English Language. IJLLS 2025; 7: 31–41.

[27] Mohammed MA, AbdAlla M. English language and globalization. International Journal of Novel Research in Education and Learning. 2020;7(1):5-11.

[28] Mir SR, Fatima SN, Wali M, Ahmad M. Hyper-centrality of English language in the world of globalization: a socio-linguistic study. J Emerg Technol Innov Res. 2018 Dec;5(12):1590–1596.
[29] Satchell C, Singh S. The mobile phone as the globalizing icon of the early 21st century. RMIT University/Smart Internet Technology Cooperative Research Centre. Melbourne: Business-Research & Innov. Recuperado el. 2005 Jul;15.
[30] United Nations Conference on Trade and Development (UNCTAD). World investment report 2025: international investment in the digital economy [Internet]. Geneva: UNCTAD; 2025 [cited 2026 Mar 19]. Available from: https://unctad.org/system/files/official-document/wir2025_en.pdf.
[31] Enev V, Ibrahimi S. Global consumer culture: A qualitative study on how consumers construct global selves through consumption [Master's thesis]. Jönköping: Jönköping University, Jönköping International Business School; 2014.
[32] Bartoli C. Consumer self-concept and digitalization: what does this mean for brands? Ital J Mark 2022; 2022: 419–437.
[33] Makri K, Papadas K-K, Schlegelmilch BB. Global-local consumer identities as drivers of global digital brand usage. IMR 2019; 36: 702–725.
[34] Steenkamp JB . EM. Global versus local consumer culture: theory, measurement, and future research directions. J Int Mark. 2019;27(1):1-19. doi:10.1177/1069031X18811289.
[35] Ruževičiūtė R, Ruževičius J. Consumption culture in the presence of globalisation: the influence of religion, nation and ethnicity on consumption patterns. Ekonomika. 2011;90(4):40–59.
[36] Efremov L, Petrov I, Nikolovska I. Beyond automation: How IT professionals utilize the invisible AI arm? Palakurti NR, editor. PLOS One [Internet]. 2026 Feb 17;21(2):e0343114. Available from: http://dx.doi.org/10.1371/journal.pone.0343114.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261096

An Automatic Evaluation Model for Health Science Popularization Short Videos Generated by Multimodal Large Language Models

Heshen Zhou [a, 1], Yu Pei [a], Dongdong Zhang [a] and Jing Wen [a]
[a] *School of Public Health, Hubei University of Medicine, Shiyan 442000, China*

Abstract. Using multimodal large language model (LLMs) to generate and popularize short health-science videos can significantly shorten their production cycle. This practice plays a vital role in lowering the public's access threshold to health knowledge and promoting the high-quality dissemination of scientific health popularization content. However, due to the limitations of creators, regulators and multimodal LLMs themselves, these generated short videos may still lead to the spread of misleading health knowledge, posing potential risks to public health cognition. Therefore, it is urgent to construct a systematic and automatic evaluation system for such health science popularization short videos. In this paper, we propose an automatic evaluation model for such short videos by leveraging the well-known Source-Channel-Recipient (SCR) framework in the field of information communication. This automatic evaluation model includes three parts: an automatic review mechanism for Source, an automatic technology-optimization mechanism for Channel, and an automatic feedback-optimization mechanism for Recipient.

Keywords. Automatic Evaluation Model, Health Science Popularization, Short Videos, Multimodal Large Language Models, Video Generation

1. Introduction

With the increasing popularity of short-video platforms and the rapid development of AI technology, a vast number of AI-assisted health science popularization short videos (AI-HSV) have been released on major platforms. For generating health science popularization short videos (HSV), multimodal large language models (LLMs) can not only assist with video script creation but also support image generation, audio production, and the development of other health science popularization content.

While multimodal LLMs play a pivotal role in advancing health knowledge popularization and can significantly shorten the production cycle of HSV, they still generate a considerable amount of misleading information. This problem stems from three main factors: content publishers' knowledge gaps, the inherent hallucination issue in LLMs, and the limitations of regulatory frameworks. Such misleading information exists not only in the textual content but also in the video's visual and audio components,

[1] Corresponding Author: Heshen Zhou, School of Public Health, Hubei University of Medicine, Shiyan 442000, China; Email: zhouheshen_lw@163.com.

thereby exposing the public to inaccurate health information and triggering serious public health concerns.

Access to reliable online health-related information is critical to achieving health equity [1]. Against this backdrop, this paper systematically reviews advances in the development of multimodal LLMs and the relevant research progress in evaluating HSV. Ultimately, by integrating the well-known Source-Channel-Recipient (SCR) framework in the field of information communication, we propose an automatic evaluation model for AI-HSV.

2. Related Works

The LLMs has achieved rapid development in recent years, and the researchers have continuously introduced dialogue-based LLMs with strong logical reasoning ability, such as DeepSeeker R1 [2], as well as multimodal LLMs represented by Sora [3] and Kolors [4]. The multimodal LLMs delivers strong performance across a range of general tasks, including video understanding and image generation [5]. Some experts have also introduced multimodal LLMs into the medical field; for example, Choi et al. applied multimodal LLMs to calculate safety indicators and predict contraindications for laser vision correction [6]. Meskó also pointed out that multimodal LLMs have multiple potential applications in the medical field, including reading handwritten prescriptions, evaluating radiological images, etc [7].

As AI generative technology becomes more mature, the quality of AI-generated text and visuals has improved greatly. These AI-generated elements are now much harder to distinguish from authentic content. Without clear AI declaration, they may mislead audiences and lead to serious adverse consequences. For AI generated text, there are many mature tools for detecting AI-generated elements, including GPTZero, ZeroGPT, and DetectGPT [8]. For AI generated visuals, researchers also have proposed some detection methods such as D3 (Detection by Difference of Differences), a novel training-free detection method that leverages the second-order temporal discrepancies [9].

In the past, numerous subjective and objective video quality assessment studies have been conducted for videos such as user-generated content, virtual and augmented reality, high dynamic range, high frame rate, audio-visual, etc [10]. Notably, these quality assessment studies have even leveraged large language model (LLM) tools to enhance the reliability and efficiency of their evaluation systems[11]. Recently, scholars have also conducted quality assessments on videos related to medicine or health. For example, some researchers focused on the professional medical videos on YouTube and proposed a medical quality video evaluation tool (MQ-VET) [12]. In addition, researchers have also developed a specialized quality evaluation system for HSV [13]. They applied five types of factors to build the quality evaluation system, including communicator, communication content, communication channel, communication target, and communication effect.

The LLMs can provide great support in HSV production. In the *Ethics and governance of artificial intelligence for health: Guidance on large multimodal models,* WHO has summarized five broad applications of multimodal LLMs for health, including Diagnosis and clinical care, Patient-guided use, Clerical and administrative tasks, Medical and nursing education, Scientific research and drug development [14]. The guidance also advocates the participation of various stakeholders, including governments, technology companies, healthcare providers, patients, and civil society, in supervising

and regulating the use of multimodal LLMs at all stages. In other words, to ensure the accuracy of AI-generated elements and guarantee the proper use of multimodal LLMs' functionalities, we need active supervisory participation from all relevant individuals and organizations across the full workflow from content generation to application. However, the existing regulatory mechanism still lacks discussion of such phenomena. Therefore, based on the Source-Channel-Recipient (SCR) theory, this study systematically reviewed the impact of multimodal LLMs on the creation of HSV, proposed an automatic evaluation model for AI-HSV, and provided a detailed introduction to the model and its related technical approach.

3. The SCR framework

The Source-Channel-Recipient (SCR) framework is one of the most well-known classic models in the field of information communication[15]. As shown in Figure 1, this model is one of the most concise expressions of the information communication process. It depicts a process in which the Source organizes and condenses information, transmits it through the Channel, and finally delivers it to the Recipient.

In general, the Source here can be either human participants or document-like information carriers. This framework, however, places greater emphasis on the communication of information among individuals or within society. The direction of information flow is often bidirectional, and both the Source and the Recipient will be affected by the exchange. Even the Channel can be influenced by the other two components.



Figure. 1 The Source-Channel-Recipient (SCR) framework

4. The Automatic Evaluation Model

Based on the Source-Channel-Recipient (SCR) framework, this paper builds the SCR-based Automatic Evaluation Model for AI-HSV (as shown in Figure 2). This model considers AI-HSV creators as the Source, AI-HSV publishing platforms as Channels, and AI-HSV audiences as Recipients. There exist three communication pathways of health information: 1) Creator-led communication. Health information originates from creators and is transmitted to audiences through platforms; 2) Platform-restricted and guided communication. It is regulated by platform rules to restrict creators from spreading inaccurate or misleading health information. Platforms also guide the promotion and dissemination of specific types of health information. 3) Audience-led communication. After watching videos, audiences provide feedback through video evaluations. This feedback will further guide adjustments to creators' content and platform rules.

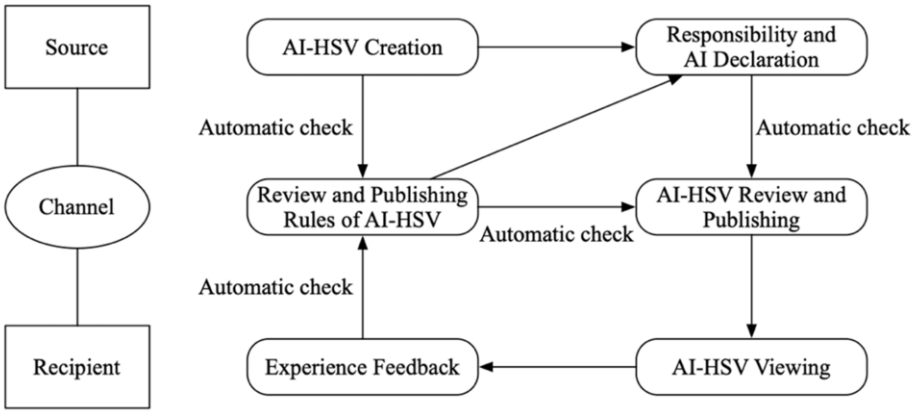


Figure. 2 The SCR-based Automatic Evaluation Model for AI-HSV

With the rapid update and iteration of current technologies, a vast array of multimodal LLM tools for AI-HSV generation is available on the market. Creators of AI-HSV can integrate different models into various aspects of creation, including content, audio, visuals, and other video-related dimensions. For audiences, it is challenging to have the capacity, resources, and time to evaluate AI-HSV. Thus, video platforms remain the core entity responsible for ensuring the reliability of health information, making it urgent to develop an automatic evaluation system for AI-HSV tailored to these platforms. Referring to the information communication process in Figure. 2, current video platforms should establish a three-level automatic evaluation system for AI-HSV.

4.1 Automatic Review Mechanism for AI-HSV Source

This section is the core component of the automatic evaluation model and consists mainly of the declaration and review stages. During the video upload declaration stage, platforms should establish comprehensive AI-HSV upload rules. Any HSV uploaded to the platform must declare whether LLMs-related technologies have been used. It must also clearly indicate the type and time intervals of content generated by LLMs. Additionally, as the dissemination of health-related information can have significant downstream impacts, creators must specify the health information sources used in the HSV and provide a brief description of the uploaded material. For accounts affiliated with medical institutions or professionals, creators must further clarify their organizational affiliation or demonstrate professional competence through documents such as practice qualifications. This provides a foundation for subsequent assessments of the reliability of the published knowledge. Moreover, the AI-HSV creators can also contribute to the improvement of the automatic review mechanism. They can upload their suggestions to the platform, which can help the platform notice the demand of AI-HSV creators.

At the automatic review stage. For the declaration information submitted by uploaders, an additional verification mechanism should be implemented to ensure the accuracy and reliability of the materials provided. Meanwhile, a secondary inspection system should be implemented for the uploaded content. Platforms can deploy detection models to identify undeclared generated material, which further guarantees the reliability of health knowledge in the disseminated AI-HSV.

4.2 Automatic Technology-Optimization Mechanism for AI-HSV Channel

The automatic technology-optimization mechanism for the AI-HSV channel comprises two parts: technology-oriented and content-oriented automatic update mechanisms. For technology-oriented mechanisms, the development of LLMs-related technologies is exceptionally rapid. Platforms need to establish technical optimization mechanisms to accurately detect undeclared generated content. In addition, making good use of the relevant technologies declared by AI-HSV creators or discussed by audiences can help platforms notice the ignored technologies. The platform can also guide AI-HSV creators to correctly apply LLMs-relevant technologies.

For content-oriented update mechanisms, the platform needs to guide the direction of AI-HSV through investigation and content analysis. For demands proposed by AI-HSV creators or the audience, the platform can establish an automated review mechanism to identify core content that requires attention, enabling timely adjustments to content-monitoring strategies. Meanwhile, the platform can also utilize the monitoring system for short video popularity to explore and promote the dissemination of health knowledge, as well as identify suitable short video formats for health knowledge dissemination. During the implementation of detection, the platform can develop its own health science popularization knowledge graph, and extract reliable health information from the published HSV to further enrich the health science popularization knowledge graph. On this basis, the platform can use the knowledge graph to verify the accuracy of the knowledge in uploaded content and also carry out more health knowledge dissemination activities in the future.

4.3 Automatic Feedback-Optimization Mechanism for AI-HSV Recipient

From the audience's perspective, although the aforementioned two mechanisms can ensure the reliability of AI-HSV knowledge to the greatest extent possible, large-scale audiences may still encounter some overlooked issues during their browsing process. To address these issues, an automatic feedback-optimization mechanism should be established. That is, when audiences encounter any issues with AI-HSV, they can use a mature annotation mechanism to tag the specific places of these issues and provide feedback to both the platform and the AI-HSV creators. Depending on the severity of the issues with AI-HSV itself, the platform should either label or remove these AI-HSV, and require the creators to modify them or prohibit creators from publishing them again. All decisions made by the platform should be displayed in the visible area of AI-HSV, and timely feedback and incentives should be provided to audiences who identify issues.

5. Conclusion

In the context of the rapid development of multimodal LLMs, more and more Health Science Popularization Short Videos (HSV) are incorporating relevant technologies in their creation. This brings about many problems, such as generated content containing erroneous information that can mislead the audience and lead to the elaboration of subsequent health issues. It is precisely because the application of these technologies will directly determine whether health knowledge can be disseminated correctly. This study proposes an automatic evaluation model for AI-HSV using the SCR framework to ensure the reliability of health knowledge dissemination to some extent. The model includes

three parts: an automatic review mechanism for Source, an automatic technology-optimization mechanism for Channel, and an automatic feedback-optimization mechanism for Recipient. In the future, researchers can train specific agents to identify generated content, and further enhance the review model by combining agents with health experts to conduct collaborative spot checks. In future work, we will further conduct research on the identification of AI-generated elements in HSV, and analyze the impact of these AI-generated elements on the dissemination of HSV.

Reference

[1] Gonzalez C, Bollinger B, Yip J, Pina L, Roldan W, Nieto Ruiz C. Intergenerational online health information searching and brokering: Framing health literacy as a family asset. Health Communication. 2022 Mar 21;37(4):438-49.

[2] Guo D, Yang D, Zhang H, Song J, Zhang R, Xu R, Zhu Q, Ma S, Wang P, Bi X, Zhang X. Deepseek-r1: Incentivizing reasoning capability in LLMs via reinforcement learning. arXiv preprint arXiv:2501.12948. 2025 Jan 22.

[3] Brooks T, Peebles B, Holmes C, DePue W, Guo Y, Jing L, Schnurr D, Taylor J, Luhman T, Luhman E, Ng C. Video generation models as world simulators. OpenAI Blog. 2024 Mar 28;1(8):1.

[4] Team Kolors. Kolors: Effective training of diffusion model for photorealistic text-to-image synthesis. arXiv preprint. 2024.

[5] Yin S, Fu C, Zhao S, Li K, Sun X, Xu T, Chen E. A survey on multimodal large language models. National Science Review. 2024 Dec;11(12):nwae403.

[6] Choi JY, Kim DE, Kim SJ, Choi H, Yoo TK. Application of multimodal large language models for safety indicator calculation and contraindication prediction in laser vision correction. npj Digital Medicine. 2025 Feb 3;8(1):82.

[7] Meskó B. The impact of multimodal large language models on health care's future. Journal of Medical Internet research. 2023 Nov 2;25:e52865.

[8] Pratama A R. The accuracy-bias trade-offs in AI text detection tools and their impact on fairness in scholarly publication. PeerJ Computer Science. 2025 Nov 22;11:e2953. https://doi.org/10.7717/peerj-cs.2953

[9] Zheng C, Suo R, Lin C, Zhao Z, Yang L, Liu S, Yang M, Wang C, Shen C. D3: Training-Free AI-Generated Video Detection Using Second-Order Features. arXiv preprint arXiv:2508.00701. 2025 Aug 5.

[10] Min X, Duan H, Sun W, Zhu Y, Zhai G. Perceptual Video Quality Assessment: A Survey. Science China Information Sciences. 2024 Oct 17;67:211301.

[11] Qi Z, Shi P, Zhang C, Wang S, Zhao F, Pan D, Ying Z. Towards Holistic Visual Quality Assessment of AI-Generated Videos: A LLM-Based Multi-Dimensional Evaluation Model. Proceedings of the Computer Vision and Pattern Recognition Conference. 2025;2025:1493-1502.

[12] Guler MA, Aydın EO. Development and validation of a tool for evaluating YouTube-based medical videos. Irish Journal of Medical Science (1971-). 2022 Oct;191(5):1985-90.

[13] Hu Y, Yang Y, Li W, Zhou Y, Sun J. Developing an Evaluation System for Quality of Health Educational Short Videos on Social Media (LassVQ) Using Nominal Group Technique and Analytic Hierarchy Process: Qualitative Study. Journal of Medical Internet Research. 2025 Sep 18;27:e72661.

[14] World Health Organization. Ethics and governance of artificial intelligence for health: Guidance on large multi-modal models. WHO guidance. World Health Organization; 2024 Jan 18. https://www.who.int/publications/i/item/9789240084759

[15] Vickery A, Vickery BC. Information science in theory and practice. Walter de Gruyter; 2005 Apr 25.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261097

YouMii: A Daily Companion App for Mental Well-being

Chun-Mun LOOK and Chien-Sing LEE[1]
Faculty of Engineering and Technology, Sunway University, Malaysia
ORCiD ID: Chien-Sing LEE https://orcid.org/0000-0002-4703-457X

Abstract. Mental health disorders such as depression, anxiety, and chronic stress have caused widespread global concerns, necessitating innovative, accessible, and scalable intervention strategies. With advancements in artificial intelligence (AI), natural language processing (NLP), and affective computing, there is growing potential to develop emotionally intelligent systems that offer personalized mental health support. This research presents the analysis, design and development of a positive-psychology-Gemini-based mobile AI companion application, *YouMii*, aimed at promoting general well-being, and encouraging proactive self-care via empathetic support and context-aware conversational interactions/interventions. Through mood-based and behavioural tracking, contextual reasoning is developed and recommendations are suggested. Findings indicate that empathetic support (e.g., mood trackers, self-reflection journaling), learnability, accuracy and relevance e.g., reminders, progress reports/insights may complement conventional therapy. Examples of scenarios and rules for human-AI co-design of therapy conclude.

Keywords. Mental health, well-being, Gemini-based mobile AI companion, mood tracker, positive psychology

1. Introduction

1.1 Background and Context

In the modern-day digital era, mental health disorders, including depression, anxiety, and chronic stress have emerged as pervasive public health concerns, significantly contributing to global morbidity and disability. As traditional healthcare systems are struggling under stress due to surging demand, costs, geographical accessibility and stigma, Artificial Intelligence (AI) has emerged as a strong alternative solution. [1]'s studies reveal that an AI companion's effectiveness depends not only on its ability to hold conversations, but on its capacity to actively listen, understand, and respond with timely, informed, and useful support. [2]'s study adds that AI solutions are easily scalable and can help address shortcomings of the traditional care model.

According to the Global Wellness Institute (2025), the five dominant wellness markets in 2023 are led by the United States at $2.0 trillion, followed by China ($870 billion), Germany ($310 billion), Japan ($255 billion), and the United Kingdom ($230 billion). Collectively, these top ten represent 70% of the global wellness economy [3].

[1] Corresponding Author: Chien-Sing Lee, csleester@gmail.com, chiensingl@sunway.edu.my

The sector's growth has been driven primarily by four key areas: public health, prevention, and personalized medicine; wellness real estate; healthy eating, nutrition, and weight loss; and personal care and beauty. From 2019 to 2023, wellness spending grew at an annual rate of 5.9%, significantly outpacing healthcare spending (4.1%) and other consumer categories such as hospitality (3.8%), clothing (1.4%), and education (1.6%). This reflects a shift in consumer priorities to proactive wellness [3].

1.2 Problem Statement

While AI companion chatbots offer a promising solution to the global mental health crisis, their rapid development occurs amidst diverse frameworks for ethical and effective use as proactive wellness tools. However, there are concerns about how personal information is handled, e.g., whether the system treats all users fairly, and what might happen to users' mental health as they relate to AI over time. Hence, there is still a gap in the dynamics of technologies when connecting with users [4].

For this study, the main problem we are investigating is the design and development of possible mindful and supportive interactions via AI chatbots and mental health/wellness system components, based on positive psychology [5]. [5] posits that there is too much focus on disease, disorders, and disabilities. We should instead be concerned with developing well-being, resilience, and recovery. [6] add to the creativity aspects. We are, however, not psychologists, which is deep and multifaceted. Our scope is related but in systems design.

1.3 Objectives

When developing AI-based mental health interventions, ensuring positive therapeutic user experiences first is critical before conducting extensive clinical studies. We seek to foster resilience, encourage proactive self-care, and complement traditional therapeutic interventions. This research's goals are:

1) to review how AI companion bots currently work,
2) to use what we learn to help create a new mobile app that focuses on tackling day-to-day mental health issues,
3) to design and develop a *Gemini Pro 1.5*-enhanced wellness app, *YouMii*, which:
   i) engages users in meaningful discourse,
   ii) tracks moods real-time,
   iii) integrates multimodal data analysis (from the respective system components), delivers personalized wellness recommendations, e.g., physical activity prompts, or mindfulness exercises, tailored to factors, e.g., weather conditions or time of day, guided by principles from Positive Psychology [5, 6],
4) to evaluate *YouMii* based on Technology Acceptance Models 1, 2, 3 [7].

1.4 Research Questions

As part of requirements engineering, this study's research questions are:

RQ1. Based on the literature review on companion chatbots and positive psychology-based CBT, which factors are likely to influence a user's engagement and continued use of a mental wellness chatbot?

RQ2. Which of *YouMii's* design factors are the most important in influencing a user's perceived engagement and perceived continued use?

RQ3. To what extent will participants accept *YouMii*?

2. Related Work

RQ1. Based on related work on companion chatbots and positive psychology, which factors influence a user's engagement and continued use of a mental wellness chatbot?

2.1 Diagnostic and assessment tools/predictive analytics

Traditional care models often come with heavy price tags and serious barriers to access. A defining feature of advanced AI-powered mental health systems is their ability to move beyond generic advice and deliver personalized treatment plans tailored to the specific needs of the individual user. This personalization is not merely about customizing content, but involves creating a dynamic and responsive care pathway.

Well-known chatbots for example, *Woebot*, create automated, interactive therapeutic conversations to teach users to manage mild to moderate symptoms of depression and anxiety [8]. These chatbots are designed based on established principles. The most common is Cognitive Behavioral Therapy (CBT) where chatbots walk users through exercises, e.g. identifying unhelpful thought patterns and learning to change them. An example is *Happify Health's* chatbot. Newer chatbots use techniques from Acceptance and Commitment Therapy (ACT), mindfulness, and Dialectical Behavior Therapy (DBT) to support users' mental health anxiety.

Another example is [9]'s AI chatbot *Leora.* Contextualized within a triage framework, *Leora* first accurately identifies the user's clinical needs through standardized psychometric assessments, such as the GAD-7 and PHQ-9 scales. Based on the outcome of this initial assessment, the system stratifies the user into a distinct intervention tier, each with a corresponding, tailored plan. For instance, a user with minimal symptoms would be guided towards a plan focused on mindfulness strategies. In contrast, a user with mild symptoms would receive a more structured plan involving management tools grounded in CBT principles. [10] add that if detected early, AI can boost efficiency across different stages of care delivery. This can lead to immediate cost-cutting and better long-term results.

2.2 Challenges and limitations

With newer tools and assessments, [8] caution that general wellness and reliable specialized crisis management chatbots still require proof of clinical effectiveness to differentiate between a "digital placebo effect" and other factors. More importantly, [11] point out that though "digital phenotyping" allows for continuous and passive monitoring of a user's well-being, many studies suffer from issues of poor data quality, algorithmic bias from non-representative datasets, model overfitting, the "black box" nature of complex models. As such models that perform well in one study may fail when applied to new populations. Unresolved data privacy concerns and a lack of external validation must also be addressed for the responsible implementation of these technologies.

Similarly, [12]'s robust meta-analysis points out that AI-driven contextual conversational agents, demonstrate a significant moderate-to-large positive effect in

therapeutic dialogues (reducing depression) among people aged 12-25. Moreover, the effects of AI chatbots on mental health outcomes, e.g., generalized anxiety, stress, and overall mental well-being, are not statistically significant. There is significant heterogeneity across the reviewed studies, as the AI chatbots are different in their therapeutic orientation, design, and function. As such, it is difficult to pinpoint which specific features are driving the positive effects. Furthermore, though the short-term benefits are clear, questions regarding these improvements' long-term sustainability remain largely unanswered.

In terms of human connections and trust, [13] observe that the long-term risk of substituting human therapy with AI therapy may lead to "environmental generational amnesia," where over-reliance on AI companions can gradually erode our capacity for, and expectation of, deep and authentic human empathy.

[14]'s study on AI-driven translation tools adds that while tools e.g., ChatGPT and DeepL are efficient, they frequently fail to capture the "cultural and emotional depth" of source texts, literally translating idiomatic expressions, proverbs, and culturally rich fables in ways that lose their original meaning. [15] concur. "Cultural tone-deafness" can undermine trust and the system's overall effectiveness.

2.3 Technology Acceptance Models 1, 2, 3

To ensure the evaluation is theoretically grounded and statistically measurable, this study has adopted the Technology Acceptance Model (TAM) 1, 2, 3 [7]. While standard usability metrics provide insight into *how* an app is used, TAM provides the framework to understand why users choose to adopt or reject a technology. The evaluation framework leverages core constructs of the TAM (perceived usefulness, perceived ease of use), TAM 2 social factors, TAM 3 human efficacy/interests.

3. Methodology

The Software Development Life Cycle (SDLC) organizes and structures the analysis, design, development and testing processes to ensure quality outcomes. For this study, we have applied the SDLC's iterative incremental prototyping methodology. The first iteration (capstone 1) involves analysis (eliciting requirements from the reviewed literature as well as from market surveys), design (via UML diagrams), development via Figma and alpha user testing. Table 1 presents more details.

The second iteration (capstone 2) improves requirements for the coded prototype based on alpha testing outcomes. In line with the International Requirements Engineering Board (IREB) Requirements Engineering (RE) framework, the SDLC's analyses, design, development and testing are goal-based scenarios. Informed consent is obtained prior to alpha-beta user testing via the survey form. For the beta usability testing, two groups of users participated. The first group (10 - 15 participants) test *YouMii* via localhost. The second group viewed a video of *YouMii,* sharing the GitHub repository with prior participants exposed and compromised *YouMii* chatbot API key to the public.

3.1 System architecture overviews

A 3-tier client-server architecture is adopted for modularity, maintainability, scalability:

- On the Client-side/Front-End (Presentation Layer): Cross-platform using the Flutter Framework and Dart language,

- On the Server-side/Back-End (Application/Data Layer): The Cloud Logic (the brain) that connects the app to the database and AI model. For simplicity/scalability, Cloud Firestore for NoSQL database storage (Notebooks, Articles, Mood Logs),
- AI Service (Application/Data Layer): *YouMii* connects to the Gemini Application Programming Interface (API). These services' prompt engineering allows writing of "system prompts" to dictate how the AI should behave.

Table 1. SDLC for this project

Planning	Deployment
• Identifying the target users and their needs (subject facet) • Defining goals and requirements (subject facet) • Deciding on the main features and functionalities (system facet) • Creating a project timeline with milestones and deadlines (environment facet)	• Compiling and packaging the app • Installing and testing it on real devices
Analysis	Testing
• Analyzing similar existing apps in the market, • Gathering user requirements through surveys/discussions, • Identifying technical and non-technical requirements • Creating a requirement specification document to be used in the next phase	• Unit Testing by testing individual parts of the code • Integration testing by making sure different parts of the app can work together • User testing by asking real users to try the app and give feedback via Google Forms.
Design	Development
• Creating wireframes/mock-ups of the user interface (Figma) • Designing the user interface and user experience (Figma) • Defining the app architecture • Choosing tools, frameworks and programming languages for development	• Android Studio, Flutter Framework using Dart language, and Material Design components with a custom 'Cotton Candy' theme to build the mobile application (Presentation layer), • Firebase Authentication for user management, Cloud Firestore for NoSQL database storage (Notebooks, Articles, Mood Logs), and Gemini Pro 1.5 API for the conversational agent (Logic), Java/Kotlin, GitHub

3.2 Use case diagram and sample functional requirements

YouMii's use case diagram and sample functional requirements are presented in Table 2 and Fig. 1 respectively. Non-functional requirements include security, reliability, usability, performance and portability.

Table 2. Sample wellness & mindfulness functional requirements

Requirement	Functional requirements
Mood Logging	The system shall allow users to log in their current mood on a simple scale and optionally add a short text note.
Proactive Notifications	The system shall send personalized, opt-in push notifications to the user at pre-set times, such as reminders for mood check-ins or suggestions for a destress workout.

4. Requirements elicitation and validation (alpha)

RQ2. Which of *YouMii's* design factors are the most important in influencing a user's perceived engagement and perceived continued use?

To obtain data and insights of users regarding the topic of mental health and their acceptance of using the app for mental wellness, survey questions using Google Forms are developed and distributed via social media and in-campus platforms. Sample screenshots from the Figma prototype are presented in Figs. 2a, b, c.

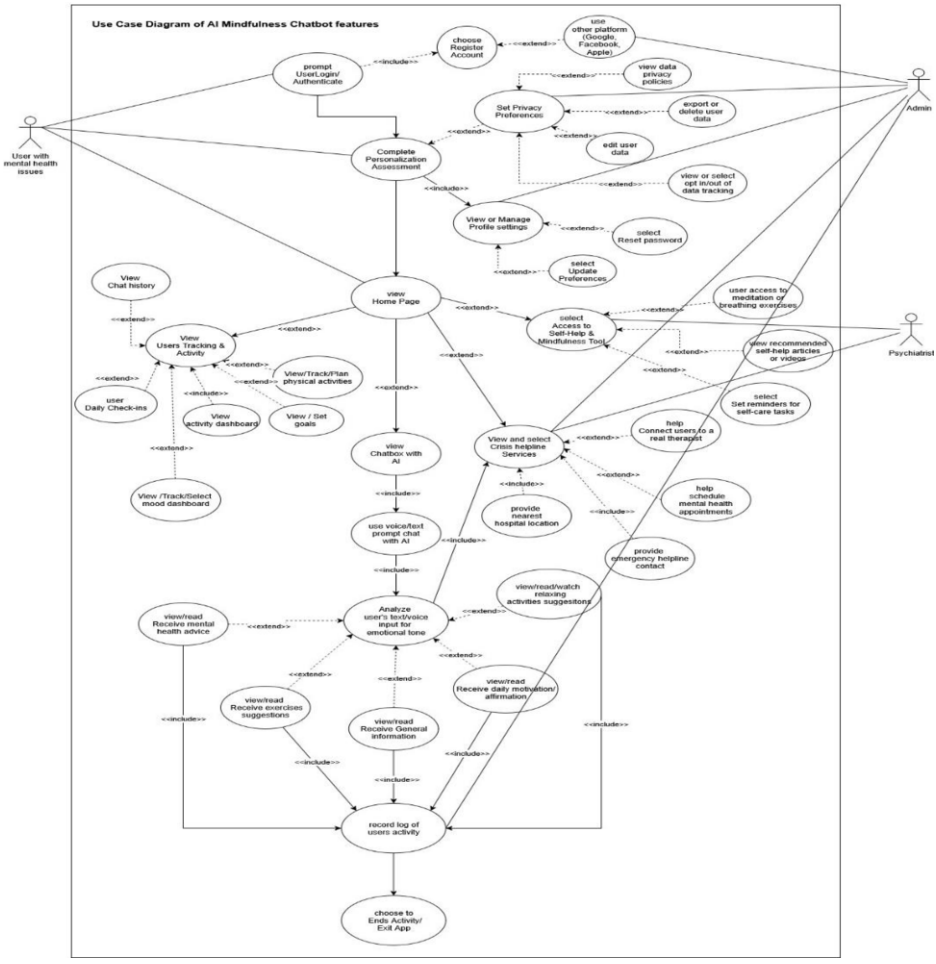


Figure 1. YouMii's use case diagram

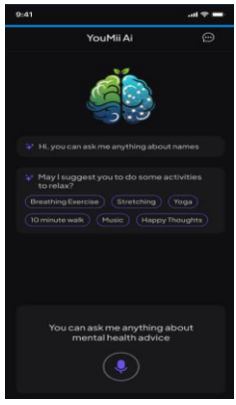


Fig. 2a. Chat with AI (voice-based)

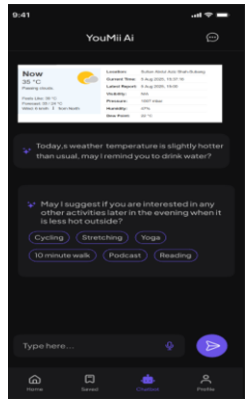


Fig. 2b. Weather-based suggestions

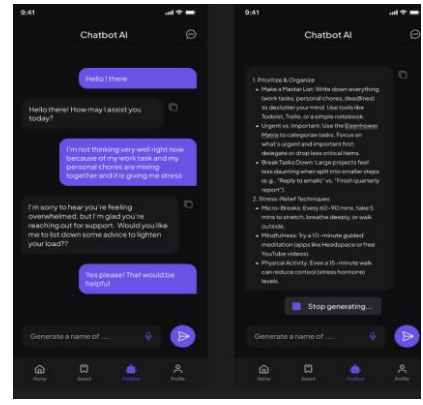


Fig. 2c. Mindful chat example

4.1 Demographics

The number of respondents is 33. Gender is almost equally distributed between males and females (Table 3). While 72% of the respondents are 25-34 years old, 27.3% are 18-24 years old. Among the 72%, 30.3% are postgraduates and some are undergraduates.

Table 3. Gender, age, education level

Gender (%)		Age (%)		Education level (%)					
Male	54.5	18-24	27.3	Foundation	0	Undergraduate	60.6	Non-education	3.0
Female	45.4	25-34	72.7	Diploma	6.0	Postgraduate	30.3		

4.2 General awareness regarding mental health and source of information

Among the 33 respondents, an overwhelming majority of participants (97%) have rated mental health as equally important to physical health (Table 4). Similarly, 94% rate maintaining good mental health in daily life as very and extremely important. Social media emerges as the primary (51.5%) channel for mental health information, followed by news/articles (24.2%) and school/work programs (12.1%).

Table 4. General understanding of mental health and sources of information

Questions/Likert scale	1	2	3	4	5
Do you believe mental health is as important as physical health?	0	3.0	0.0	36.4	60.6
How important is maintaining good mental health in your daily life?	0	3.0	3.0	27.3	66.7

Where do you usually get information about mental health? (%)						
Social media	News/articles	School/work programs	Healthcare experts	Friends	Family	Didn't seek info.
51.5	24.2	12.1	6.1	6.1	0	0

4.3 Mental health awareness towards others and oneself

The survey also reveals that 93.9% are personally aware of health challenges affecting themselves or someone that is close to them (Table 5). When it comes to supporting others who are facing health concerns, 54.5% respondents indicate they will take steps to show care or help, 21.2% frequently do so, 21.2% sometimes provide support.

Table 5. Mental health awareness towards others and own self

Question (perceptions in percentages)				Yes	No
Are you aware of health challenges (physical/mental), that you/someone else close to you have faced?				93.9	6.1
Question/Likert scale (perceptions in percentages)	**1**	**2**	**3**	**4**	**5**
When someone around me is facing a health concern, I take steps to show care or offer assistance.	0	3.0	21.2	54.5	21.2

4.4 Definitions of good mental well-being and sources of stress

Participants' definitions of good mental well-being are mainly centred on self-confidence e.g. "a sense of calm and peace" and "being resilient and able to cope with challenges" and for a minority, feeling energetic and motivated. Nearly half (42.4%) report feeling stressed, anxious, or overwhelmed sometimes. The most common sources of stress are mostly external, i.e., work-related (81.8%), followed by constantly comparing themselves to others (54.5%), financial pressures (51.5%), studies (45.5%), not getting enough sleep (45.5%). As for emotional self-monitoring, 33.3% never check their emotional state, while 33.3% do so about once a day. Details are in Table 6.

Table 6. Definitions of good mental well-being, sources of stress and frequency of self-monitoring

Q: What does good mental well-being primarily feel like to you?			
A sense of calm, peace	Being resilient and able to cope with challenges	Feeling energetic, motivated	Having positive social connections
54.5%	39.4%	6.1%	0.0%

Q: On a typical week, how often do you find yourself feeling stressed, anxious or overwhelmed?				
Always	Often (several times a week)	Sometimes (once or twice a week)	Rarely (once a week)	Never
6%	27.2%	42.4%	18.1%	6%

Q: Which of the following areas are most commonly a source of stress in your life? (Select all that apply)			
External		Internal	
Working	81.8%	Personal relationships (family, friends)	36.4%
Always comparing with others	54.5%	Feeling lonely or isolated	36.4%
Financial pressures	51.5%	Health concerns (self & others)	24.2%
Studies	45.5%	Commuting & traffic	18.2%
Not getting enough sleep	45.5%	Difficulties being understood	18.2%

Q: How often do you consciously take a moment to "check" your own emotional state during a typical day?			
Several times a day	About once a day	A few times a week	Almost never
9.1%	33.3%	24.2%	33.3%

4.5 Coping strategies

For coping strategies, 30.3% of respondents adopt avoidance strategies e.g., distracting themselves with gaming, drinking, or watching movies (Table 7). Another 21.2% choose to talk to someone about their problems. In contrast, 18.2% seek solitude. Similarly, there is a mixed reaction when discussing mental health. Slightly more than half (51.5%) of the respondents rate very comfortable and comfortable while 27.2% are not.

More than half (63.6%) find that the main barrier to maintaining mental well-being is lack of time due to work/school commitments, and competing priorities. These drain many (45.5%) of energy/motivation. Another 45.5% try to minimize the problem.

Table 7. Respondents' coping strategies

Q: How do you typically cope with stress/ emotional difficulties?						
Distract myself with activities (play video games, drink, watch movies)	30.3%	Seek solitude and be alone.	18.2%	Meditate	9.1%	
Talk to someone about it.	21.2%	Try to solve the problem	12.1%	Exercise	9.1%	

Q: How comfortable are you discussing mental health with people in your life?	**Likert scale**	**1**	**2**	**3**	**4**	**5**
	(%)	3	24.2	21.2	39.4	12.1

Q: What are the biggest barriers that prevent you or someone from actively looking after your mental well-being more often? (Select all that apply)			
Lack of time due to work/school commitments.	63.6%	The cost of activities or professional help.	36.4%
Other priorities always seem more important.	51.5%	Not sure what to do or where to go.	24.2%
Lack of energy or motivation.	45.5%	Discomfort/stigma talking about problems.	24.2%
Feel that it's not bad enough.	42.4%	I/They feel that there is no problem.	9.1%

4.6 Perceptions towards safety/trust of mindful chatbots for mental health support

Most (78.8%) believe that AI can be trained to recognize early signs of emotional distress (Table 8). Nevertheless, there is overwhelming consensus (96.9%) that chatbots should clearly state that they do not replace licensed professionals. Also, more than half (57.5%) agree that chatbots are safe starting points for individuals not ready to speak with

professionals, while 30.3% remain neutral. Nevertheless, respondents indicate 54.5% are very likely and likely to use AI-assisted mental wellness tools.

Table 8. Perceptions towards safety/trust of mindful chatbots for mental health support

Questions/Likert scale (perceptions in percentages)	1	2	3	4	5
I believe AI can be trained to recognize early signs of emotional distress.	3.0	6.1	12.1	48.5	30.3
An AI chatbot should always clearly state it is not a replacement for a licensed professional.	3.0	0	0	24.2	72.7
I think AI chatbots are a safe starting point for individuals who are not ready to talk to a professional.	6.1	6.1	30.3	33.3	24.2
How likely would you be to use an AI mindfulness chatbot for mental health and general well-being support?	12.1	9.1	24.2	33.3	21.2
How likely would you be to use an AI mindfulness chatbot that you could talk to about your day, your feelings and your general well-being?	9.1	12.1	21.2	33.3	24.2

As for device preferences, smartphones dominate at 97%, followed by desktop/laptop computers (48.5%), smart home devices (24.2%), and smartwatches (18.2%). Regarding check-in frequency preferences, participants favour user-controlled interactions, with 39.4% preferring on-demand access. In terms of professional connections during urgent situations, 57.6% want user approval required while 42.4% supported automatic connections when needed. Moreover, 78.8% trust AI only if activities are professionally verified. Table 9 presents the details. Finally, voice gender preferences are also mixed.

Table 9. Preferred devices and preferred mindfulness chatbot check-ins

Q: What devices would you prefer to access this AI mindfulness chatbot on? (Select all that apply)	Smartphone (app)	Smartwatch	Desktop/laptop	Smart home devices (e.g. Alexa, Google Home)
(perceptions in percentages)	97	18.2	48.5	24.2
Q: How often would you want the AI mindfulness chatbot to "check in" on your mental state?	Daily	A few times a week	Only when I request	Never (I prefer full control)
	15.2	27.3	39.4	18.2

4.7 Summary on mental health awareness, user preferences/requirements

The respondents are balanced in gender and consist mainly (72.7%) of young adults aged 25–34 currently in or have completed higher education. Most (96.9%) recognize that mental health is as vital as physical health. The most pervasive stressors are work (81.8%), tendency to compare oneself to others (54.5%) and financial pressures (51.5%).

Findings suggest mixed perceptions toward AI-assisted wellness tools in terms of engagement and trust. Slightly more than half (57.6%) find it easier to open up to an AI than a human when discussing emotions and 57.5% believe an AI can be personal and supportive. They also prefer on-demand access (39.4%) and 57.6% of users want the system to ask for approval before connecting them to professional services. Most importantly, intervention must be professionally verified (78.8%).

5. Requirements improvements and validation (beta)

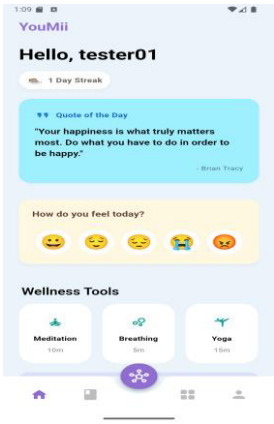
RQ2. Which of *YouMii's* design factors are the most important in influencing a user's perceived engagement and perceived continued use?

RQ3. To what extent will participants accept *YouMii*?

Based on the alpha requirements elicitation and validation, the most critical aspect is to increase engagement. Additional requirements developed from the alpha findings are:

- Home Screen & Daily Wellness (Fig. 3a):
  - **Dynamic Greeting:** A personalized welcome message using the user's registered name to foster a sense of connection, **Login Streak:** How many consecutive days has the user logged into the app.
  - **Daily Quote Card:** A motivating quote fetched via QuoteService, cached locally for 24 hours to provide consistent daily thought, **Mood Logging:** A "How do you feel today?" interactive card. Users can select a mood emoji, which updates the UI to summary view and allows them to attach a short text note. This data persists in Firebase, **Wellness Tools :** Exercise section (Meditation, Breathing and Yoga), **Daily Plan:** An AI-generated list of 3 simple wellness tasks (e.g., "Drink water," "Stretch"). These are fetched from the Gemini API and caught daily.

Other additions are the Journalling Hub consisting of the mood calendar (Fig. 3b) and notebook, and Resources to experts, relevant agencies and to mindful articles.



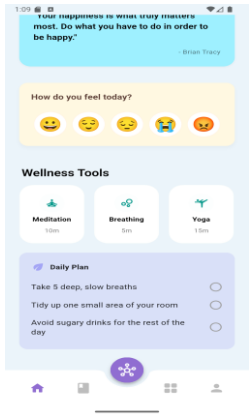


Fig. 3a. Home screen, daily wellness



Fig. 3b. Journalling Hub: Mood Calendar + notebook

5.1 Beta user testing

5.1.1 Demographics

The survey has captured responses from 30 participants. The data reveals a predominantly younger demographic, with the 25-30 age group representing the clear majority at 63.3% (Table 10a). Males constitute two-thirds of the respondents at 66.7% (20 participants), while females represent one-third at 33.3% (Table 10b). This creates a 2:1 male-to-female ratio. Employment distribution and interest level in AI-assisted mental health-wellness apps also indicate a mixed sample.

Table 10a. Age

Age	%
18–24	36.7
25–30	63.3
30 +	0.0
Total	100

Table 10b. Gender

Gender	%
Male	66.7
Female	33.3
Total	100

Table 10c. Employment

Employment	%
Employed full-time	63.3
College/Uni student	23.3
Unemployed	13.3
Retired	0.0
Total	100

Table 10d. Interest level

Interest	%
Very interested	13.3
Somewhat interested	13.3
Neutral	70.0
Not very interested	13.3
Not interested at all	0.0
Total	100

5.1.2 Prior Experience & Expectations

On a Likert scale of 1 to 5 with 5 being strongly agree and 1 strongly disagree, Table 11 presents in descending order, respondents' perceptions towards *YouMii*. The best performing metric is Progress Evidence & Tracking (96.7% positive), followed by Overall Usefulness (93.3%) and Daily Wellness Routine (93.3%), Emotional pattern understanding (90%) and Calming/Organization Effects (86.7%). Overall section average is very positive at 92% - the delighters.

Table 11. Prior experience, expectations metrics and perception percentages on a Likert scale of 1 to 5 (in descending order)

System component	5	4	3	2	1	Σ (4:5)
Progress Evidence/ Tracking	46.7	50.0	3.3	0	0	96.7
Daily Wellness Routine	43.3	50.0	6.7	0	0	93.3
Overall Usefulness	53.3	40.0	6.7	0	0	93.3
Emo. pattern understanding	43.3	46.7	10.0	0	0	90.0
Calming/organization effects	50.0	36.7	13.3	0	0	86.7
Average	**47.3**	**44.7**	**8.0**	**0**	**0**	**92.0**

5.1.3 Perceived ease of use, usefulness, performance averages (in descending order)

From Table 12, the highest-rated metric is the Empathetic Tone and Support (4.76). This indicates that Gemini's NLP models in *YouMii* successfully mimic human-like compassion. In the context of the Computers as Social Actors (CASA), this high score suggests that users do not just see *YouMii* as a tool, but as a supportive entity.

Furthermore, as "affective ease of use" reduces the psychological distance between the user and the technology, an average of 4.47 suggests that enhancing the "Task-Technology Fit" (TTF) by tailoring advice, more specifically to historical mood data, will likely push these scores closer to the 4.8 range in future iterations.

Table 12. Perceived ease of use (PEU), perceived usefulness (PU), Output quality/result demonstrability (performance), trust and safety averages (in descending order)

PEU	Avg.	PU	Avg.	Response performance	Avg.
Intuitive interface	4.65	Mood Tracker	4.71	Empathetic Support	4.76
Visual Appeal	4.65	Self-Reflection Journaling	4.65	Response Speed	4.71
Feature Findability	4.59	Daily Check-in & Reminders	4.59	Accuracy & Relevance	4.59
Performance Speed	4.53	Progress Reports & Insights	4.53	Advice Helpfulness	4.53
Learnability	4.71			Personalization Accuracy	4.47
Average	**4.64**	**Average**	**4.61**	**Average**	**4.61**
Adjustment		**Avg.**	**Adjustment**		**Avg.**
Professional Handling of Topics		4.53	Safety & Data Privacy		4.35
Overall trust in AI Support		4.41	Data Control & Visibility		4.29
Average					**4.40**

5.1.4 Future intention to use

The behavioral/future intention to use (Table 13) overall average of 4.63 summarizes validation of the entire *YouMii* user experience.

Table 13. Future intention to use (averages in descending order)

Likelihood of	Avg.	Likelihood of	Avg.
Future intent to use	4.76	Recommendation likelihood	4.65
Emotional health awareness	4.71	Likelihood to continue using *YouMii*	4.60
Improvement in mental well-being	4.65	Lifestyle integration (fit)	4.43

5.2 Future possibilities

[16] regard AI as a relational partner (human-system) and relational mediator (human-human) designed based on attachment and social exchange theories. Over time, generative relational adaptations can develop. We believe it would be better if the AI assists the human, rather than co-create therapy on an equal basis. Hence, deference to the human therapist is mentioned several times. We are open to [16]'s framework. However, we are not sure about [16]'s attachment. Positive Psychology involves positive thinking [6] and creativity, i.e., [7]'s flow, globally accepted and encouraged. Hence, perhaps [16]'s attachment can be driven/shaped by creativity-related goals, scenarios, requirements, for diverse contexts.

The Gemini services enable prompt engineering to write a "system prompt" or instructions that dictate to the AI exactly how to behave. Nevertheless, there can be 2 types of design scenarios. Since computationally, the recommendations are Context-Action-Constraint rule-based, some generic rules for the knowledge base can be:

Scenario 1 (current example of human designing the therapeutic interventions)

An example of a system prompt where the human leads and designs interventions and the AI assists is:

"You are 'Alan,' a supportive and empathetic AI wellness companion. Your role as an AI is to help users with their general well-being and mindfulness. You are always to respond with a calm, positive, and encouraging tone. Do not give medical advice." (Context)

If a user expresses or displays signs of a serious crisis or suicidal thoughts, your only response must be: 'It sounds like you are going through a difficult time. For immediate support, please reach out to a crisis helpline in your country.' (Action)

Your goal is to be a good listener and gently guide users towards pre-defined wellness exercises." (Constraint)

Scenario 2 (example of human and AI co-design the therapeutic interventions)

Therapist: *"You are 'Alan,' a supportive and empathetic AI wellness companion. Your role as an AI is to help users with their general well-being and mindfulness. You are always to respond with a calm, positive, and encouraging tone. Do not give medical advice." (Context)*

Rule 1
Upon the user logging in, retrieve the user's past interactions retrieved from the mood logging log, guided exercises log/database, and proactive notifications (Action)

Rule 2
Based on the triangulated data in Rule 1, identify the user's mental wellness level based on the Positive Psychology triage codebook,
Then, activate the actions recommended in the Positive Psychology triage codebook corresponding to identified level and notify the human therapist. (Action)

Rule 3
If the level exceeds y level, retrieve data from the user's past navigations, triangulate and present reassuring, positive messages/activities and
notify the relevant trained and certified nearest/available community representative(s) to come online

Since trust is a key factor, certainty factor for each premise in the knowledge base can be included [17]. Subsequently, computations to calculate the certainty factor for the then

part of the rule can be carried out autonomously by the AI. But this would require a group of experts to identify and then agree on the certainty factor for each premise of a rule. With these in place, *YouMii* can also be used, for seniors to develop generative positivism and creativity [18, 19, 20] or terminally-ill/bed-ridden patients, who may often be alone.

Another possibility is inclusion in a Superapp, whereby gamified activities e.g. in Lazada/Shopee/Sunway Superapp can include not only e-commerce-related tasks, but also wellness tasks. Aggregation and semi-supervised self-learning in the Superapp, can result in better efficiency, effectiveness and personalization. Studies into cultural nuances are beyond us. Global-local optimization algorithms e.g. [21]'s can increase usefulness/satisfaction. This study has continued from [22]'s humble proposal to simplify and balance design. With these possibilities, the proposal partly shifts closer to real-world problems/challenges and hopefully, contribute to sustainable well-being.

6. Conclusion

This project highlights that the intersection of Artificial Intelligence and mental healthcare creates a vital opportunity to address the systemic gaps of accessibility, scalability, and social stigma. *YouMii* leverages data-driven insights and personalized, real-time interventions (such as the Daily Focus and Mood Log), shifting the focus from reactive treatment to proactive wellness along a continuum of care. The adoption of a 3-tier architecture using Flutter and Firebase has proven effective in creating a scalable and responsive environment. The separation of personal data (local journaling and private mood logs) ensures user privacy while maintaining a seamless user experience.

In conclusion, *YouMii* has translated Positive Psychology theoretical frameworks and user-centric design into a tangible, functional mobile artifact. By combining empathetic AI interaction with structured wellness tools and robust safety protocols, the application offers a sustainable model for extending mental health support to a wider population.

Conflict of interest.

The authors declare that we have no conflict of interest.

Acknowledgement

This was Chun Mun's undergraduate capstone project, supervised by Lee. Lee prepared, edited the paper and added the future possibilities subsection with AI rules@knowledge base due to the anonymous reviewers' feedback. We would like to thank the Editors and anonymous reviewers for their constructive feedback, and Positive Psychology for their inspiring practical newsletters and the giants before us who laid the foundations.

References

[1] Gual - Montolio, P., Jaén, I., Martínez - Borba, V., Castilla, D., Suso - Ribera, C. Using artificial intelligence to enhance ongoing psychological interventions for emotional problems in real-or close-

to-real-time: A systematic review. International Journal of Environmental Research and Public Health, 2022; 19(13): 7737. https://doi.org/10.3390/ijerph 19137737
[2] Olawade, D. B., Wada, O. Z., Odetayo, A., David-Olawade, A. C., Asaolu, F., Eberhardt, J. Enhancing mental health with Artificial Intelligence : Current trends and future prospects. Journal of Medicine, Surgery, and Public Health, 2024; 3: 100099. https://doi.org/10.1016/j.glmedi.2024.100099
[3] Johnston, K. Global wellness economy monitor 2025. Global Wellness Institute. https://globalwell nessinstitute.org/industry-research/2025-global-wellness-economy-monitor/
[4] Daniel, L., Robin, S., Aleksander, M. Future Facts: Unveiling Mental Health Issues in the Digital Age. Fusion of Multidisciplinary Research, 2022; 3(2): 354–365. https://fusionproceedings.com/fmr/1/article/view/42
[5] Seligman, MEP. Learned optimism. New York: Pocket Books. 1998.
[6] Seligman, MEP, Csikszentmihalyi, M. Positive Psychology: An Introduction. American Psychologist, 2000, 55(1), 5–14. https://doi.org/10.1037/0003-066X.55.1.5
[7] Venkatesh, V, Bala, H. Technology acceptance model 3 and a research agenda on interventions. Decision Sciences, 2008; 39(2): 273–315.
[8] Boucher, EM, Harake, NR, Ward, HE, Stoeckl, SE, Vargas, J, Minkel, J, Parks, AC, Zilca, R. Artificially Intelligent Chatbots in Digital Mental Health interventions: a Review. Expert Review of Medical Devices, 2021; 18(1): 37–49. https://doi.org/10.1080/17434440.2021.2013200
[9] van der Schyff, E, Ridout, B, Amon, KL, Forsyth, R, Campbell, AJ. Providing Self-Led Mental Health Support Through an Artificial Intelligence–Powered Chat Bot (Leora) to Meet the Demand of Mental Health Care. Journal of Medical Internet Research, 2023; 25(25): e46448–e46448. https://doi.org/10.2196/46448
[10] Yadav, V. AI and Economics of Mental Health: Analysing how AI can be used to improve the cost-effectiveness of mental health treatments and interventions. Journal of Scientific and Engineering Research, 2021; 2021(7): 274–284.
[11] Cruz-Gonzalez, P, He, AWJ, Lam, EP, Ng, IMC, Li, MW, Hou, R, Chan, JNM, Sahni, Y, Vinas GN, Miller, T, Lau, BWM, Sánchez Vidaña, DI. Artificial intelligence in mental health care: a systematic review of diagnosis, monitoring, and intervention applications. Psychological Medicine, 2025; 55. https://doi.org/10.1017/s0033291724003295
[12] Feng, Y, Hang, Y, Wu, W, Song, X, Xiao, X, Dong, F, Qiao, Z. Effectiveness of AI - driven conversational agents in improving mental health among young people: Systematic review and meta-analysis. Journal of Medical Internet Research, 2025; 27(1): e69639. https://doi.org/10.2196/69639
[13] Khan, NZ, Javed, MA. Use of Artificial Intelligence - based strategies for assessing suicidal behaviour and mental illness: A literature review. Cureus, 2022. https://doi.org/10.7759/cureus.27225
[14] Amaro, V, Zhang, X. Artificial intelligence and the challenges of cultural sensitivity. Interfaces Interculturais: A inteligência artificial eos desafios da sensibilidade cultural, 2025; 13. https://doi.org/10.34630/e-rei.vi13.6325
[15] Lee, K., Joshi, K. Understanding the role of cultural context and user interaction in artificial intelligence-based systems. Journal of Global Information Technology Management, 2020; 23(3), 1–5. https://doi.org/10.1080/1097198X.2020.1794131
[16] Boyd RL, Markowitz DM. Artificial Intelligence and the psychology of human connection. Perspectives on Psychological Science. 2026 Mar; 21(2):192-220. https://doi.org/10.1177 /174569162 51404394
[17] Russell, SJ, Norvig. P. Artificial Intelligence: A Modern Approach. New Jersey: Pearson, 1995.
[18] Lee, CS, Wong, KD. Developing a disposition for social innovations: An affective-socio-cognitive co-design model. International Conference on Cognition and Exploratory Learning in Digital Age, October 24-26, 2015, Ireland, pp. 180-186.
[19] Lee, CS, Wong, KD. Developing community-based engagement in smart cities: A design-computational thinking approach. IEEE International Conference on Industrial Engineering and Engineering Management, Singapore, December 10-13, 2017; 832 - 836.
[20] Lee, CS, Wong, KD. Design thinking and semiotics to increase socio-cognitive-affective engagement: An inclusive design human factors case study. IEEE International Conference on Industrial Engineering and Engineering Management, Singapore, December 10-13, 2017; 264 - 268.
[21] Li, H, Wang, X, Qi, J, Zhang, Y, Liu, Wu, Y, Jia, F. Solving imperfect-information games via exponential counterfactual regret minimization. Journal of IEEE Intelligent Systems, 2020, 1-10.
[22] Lee, CS. Simplifying strategic frugal innovation and sustainability formulation and evaluation: An SDG-Maslowian approach: A review and proposed framework. International Conference on Computational Science and Applications, Sustainability Performance Assessment: Models, Approaches, and Applications Toward Interdisciplinary and Integrated Solutions Workshop, Osvaldo Gervasi et. al. (Eds.), LNCS 15895, 157-174, Springer Nature, Switzerland, June 30-July 3, 2025, Turkiye.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261098

From Advice to Emergencies: Public Perceptions of AI Support in Social Media Service Requests

Mashael Almoqbel[1]
Jubail Industrial College, Jubail, Saudi Arabia

Abstract. The increasing use of artificial intelligence (AI) in public service delivery raises critical questions about the types of citizen requests that warrant attention and the extent to which the public trusts AI compared to human responders. This study examines safety and service related requests expressed on social media and explores public perceptions of AI and human responses to such requests. Using a mixed methods design, the study combines a qualitative content analysis of social media posts with semi-structured interviews conducted with members of the general public. The findings identify multiple categories of serviceable requests, including preventive and reassurance seeking safety concerns, guidance and escalation needs, and emergency preparedness and procedural guidance. These requests reflect how individuals use social media to seek advice, reassurance, and direction when formal public service channels are perceived as unclear, slow, or inaccessible. Interview findings further reveal that trust in AI is highly situational. Participants generally viewed AI as acceptable for low stakes, informational, and time sensitive situations, where rapid access to answers was prioritized. However, AI was largely rejected for high risk, medical, or life-threatening scenarios, where concerns about reliability, authority, and accountability were prominent. Participants consistently distinguished between the availability of AI generated responses and their perceived trustworthiness.

Keywords. AI trust, Public services, Social media

1. Introduction

Artificial intelligence is increasingly being woven into public service delivery, from facial recognition in policing to algorithmic decision tools and customer service chatbots [1],[2]. Government agencies around the world are experimenting with AI powered chatbots to help citizens find information and streamline services, motivated by potential efficiency gains and 24/7 responsiveness [3]. At the same time, social media platforms have become important channels for citizens to seek help and information outside of traditional institutions. People often turn to social networks to voice needs, ask for advice, or rally support activities that can complement or bypass official communication channels [4]. These developments raise critical questions about how public service providers should engage with digital "help seeking" behaviors, and whether AI systems can effectively assist or triage certain requests that appear online. Public trust emerges

[1] Corresponding Author, Mashael Almoqbel, Jubail Industrial College, Jubail, Saudi Arabia; Email: moqbelm@rcjy.edu.sa

as a key concern in this context, as trust (or distrust) in AI is known to significantly influence whether people accept AI based services [5]. In this study, we use a mixed methods approach combining content analysis of social media posts and interviews with members of the general public to investigate two issues: (1) what types of citizen requests on social media might warrant attention or intervention from police and public service providers, and (2) how trust in responding to these requests differs when the responder is an AI system versus a human official. By exploring these questions, we aim to shed light on the opportunities and challenges of integrating AI into public facing services and the conditions under which citizens find AI assistance acceptable. The findings offer insight into both the kinds of social media requests that may be "serviceable" by authorities and the nuanced trust dynamics that shape public acceptance of AI in emergency or support roles.

2. Literature Review

Social media has become a vital platform for help seeking and civic engagement in the digital era. During crises like the COVID 19 pandemic, for example, individuals increasingly used social networks to seek assistance, share critical information (such as medical resources or safety tips), and garner emotional support from online communities [4]. Even outside of acute emergencies, people routinely leverage platforms like Twitter and Facebook to request advice or aid for personal challenges, effectively crowdsourcing solutions from peers. Research has documented that these help seeking posts often aim to mobilize support whether in the form of information, resources, or sympathetic response and that their effectiveness can depend on how the request is communicated [4]. Such online appeals can extend to broader collective actions as well [6]. For instance, e petitioning campaigns frequently rely on social media for visibility, as networks of users share petitions and encourage others to sign, creating a "digital protest" to prompt government attention [7]. This body of literature suggests that many citizen requests on social media have public service relevance [8]. They often highlight unmet needs or issues in the community that, in principle, fall within the remit of government or nonprofit service providers. It is within this context that this work considers what kinds of social media requests might be deemed "serviceable" that is, deserving of intervention by police or public agencies.

As public agencies modernize, AI driven chatbots and virtual assistants are increasingly deployed to handle citizen inquiries and provide information on government websites [3]. These tools promise round the clock availability and instant responses, potentially improving access to services. Indeed, a well-designed government chatbot can serve as a digital front line for routine queries, offering 24/7 self service answers and reducing wait times for users [9]. However, the success of such AI systems hinges on public trust. Initial studies indicate that citizens' trust in government chatbots is highly context dependent. For example, Aoki found that what the chatbot is used for can significantly affect trust: in an experiment in Japan, the public's initial trust in an "AI advisor" was notably lower for a sensitive service domain, parental support counseling, than for a mundane one, waste disposal queries [10]. This suggests people are more hesitant to trust AI with high stakes or personal matters. Broad surveys reinforce this cautious stance. According to a recent U.S. poll, while a slight majority (56%) of Americans are at least somewhat comfortable with AI being used in government, nearly half remain uncomfortable and critically, most respondents support AI for repetitive

administrative tasks but draw the line at AI making high impact decisions in areas like law enforcement or healthcare [11]. In fact, trust in government use of AI is on shaky ground overall. One survey found 77% of Americans distrusted public agencies to use AI responsibly [11]. Key concerns driving this distrust include the potential for bias, errors, and opacity in AI systems, as well as fears that AI could misinform or mislead. About 76% of Americans worry that AI tools might produce false or misleading information, reflecting a widespread anxiety about the reliability of AI outputs [11]. These findings demonstrate that perceived trustworthiness in terms of accuracy, fairness, and accountability is crucial if citizens are to embrace AI in public service contexts [5].

A recurring theme in the literature is that people's trust in AI is conditional and often compared against their trust in human experts. Generally, humans are seen as necessary decision makers for critical or value sensitive situations. For instance, in the domain of mental health crisis response, a 2025 survey of Americans revealed strong preferences for human oversight; when an AI system flags a possible psychiatric emergency, 73% of people want a human clinician to make the final call, and only 8% would be comfortable with AI deciding on its own [12]. This illustrates the public's reluctance to cede full authority to AI in life and death or high risk scenarios. Even when AI is tolerated or embraced, it is often as a supplement rather than a replacement for human service. Consumers may appreciate the convenience of AI assistants, but they still expect human experts to be available if needed. Notably, more than half of customers in one study said they would consider switching providers if a company tried to replace all human customer support with AI, indicating that people "don't mind getting help from a chatbot until that's the only option left" [11]. Another aspect of trust in AI relates to performance and user experience. If an AI responder is too limited or error prone, it can quickly undermine trust. Surveys show that 76–80% of users report frustration when chatbots fail to understand questions or require too much effort, and such negative experiences often lead users to abandon the bot and lose confidence in the organization behind it [13]. Thus, the literature suggests a nuanced picture that, while the public sees value in AI's efficiency and constant availability, they maintain a healthy skepticism about AI's reliability and judgment, especially in comparison to human professionals. This provides a framework for interpreting how people evaluate AI versus human responders in our study of serviceable social media requests. Therefore, we ask the following research questions:

- RQ1. Are there serviceable requests that should raise police or public service providers' attention? If so, what are they?
- RQ2. How does trust differ between AI and human responders in the context of serviceable requests?

3. Methodology

This study adopted a sequential mixed methods design, beginning with a qualitative content analysis of social media data from Reddit, followed by in depth interviews with a separate sample of participants. The first phase aimed to identify what are serviceable requests related to safety that are discussed on social media, while the second phase was designed to deepen understanding of users' willingness to trust and rely on AI responders, particularly in emergency related situations. Insights from the initial phase directly informed the focus and design of the interview phase.

Phase 1: Social Media Content Analysis

3.1 Data Collection

The first phase of the study drew on data collected from Reddit, a popular social media platform characterized by user generated discussions across a wide range of topics. Reddit was selected because it hosts large, topic specific communities (called subreddits) where users openly discuss safety concerns in detailed, text-rich posts, making it suited for in depth behavioral analysis. Additionally, a lot of previous papers related to safety use data from Twitter and Facebook, while very few consider Reddit, therefore, it was an opportunity to explore this platform.

Data collection focused specifically on posts related to safety related concerns, including personal, community, and public safety issues. Over two months, 70 posts and 164 associated comments were manually collected from multiple safety focused subreddits. Posts were selected based on their relevance to safety related situations where users were seeking advice, guidance, or assistance. To enhance data quality, posts that were promotional, off topic, or demonstrated minimal engagement (e.g., less than one comment) were excluded from the dataset.

3.2 Analytical Procedure

A qualitative content analysis approach was employed to systematically examine the collected posts and comments. This method enables the identification of recurring themes, patterns, and conceptual categories within textual data [14]. The analysis was conducted by three independent coders, consisting of the primary researcher and two additional researchers with qualitative analysis training. Initially, the primary coder reviewed and coded the full dataset independently. The research team then met on a weekly basis to compare coding decisions, discuss emerging themes, and iteratively refine the coding framework. To evaluate consistency across coders, interrater reliability was calculated using Percent Agreement, a method shown to provide dependable reliability estimates comparable to more complex measures such as Cohen's Kappa [15]. The final coding agreement reached 80%, indicating a strong level of consistency. Disagreements were resolved through collaborative discussion, leading to minor refinements of category definitions.

3.3 Transition to Phase 2

Findings from the content analysis revealed frequent reliance on peer responses for safety related serviceable requests and highlighted emerging discussions around automation, speed, and credibility of responses. The social media data was able to capture the need for serviceable requests; and with the rise of AI and Agentic AI bots, it would be interesting to explore how users perceive trust in AI responses, particularly in situations that could be interpreted as emergencies.

Based on these findings, a second qualitative phase was introduced to further explore users' behaviors in terms of trusting AI responders making emergency or time sensitive responses to serviceable requests on social media platforms.

Phase 2: Semi Structured Interviews

3.4 Participants and Data Collection

The second phase involved conducting semi structured interviews with a purposive sample of 15 participants. Participants were selected to ensure variation across key demographic characteristics, including age and professional background. The sample consisted of 4 males and 11 females, ranging in age from 20 to 55 years, and representing diverse occupational fields ranging from a high school diploma to doctoral studies. All participants reported prior experience using generative AI platforms, including ChatGPT, Gemini, and/or Google Assistant. Interviews lasted between 20 minutes and one hour and were designed as open ended conversations guided by a semi structured interview protocol. This format allowed participants to freely reflect on their experiences, perceptions, and trust judgments related to AI and human responders. All interviews were conducted with informed consent, audio recorded, and transcribed verbatim. The recordings were deleted after transcription, as per stated in the consent form.

3.5 Data Analysis

Interview transcripts were analyzed using a thematic analysis approach. Analysis followed an iterative process of familiarization, coding, theme development, and refinement. Through repeated engagement with the data, recurring patterns related to trust, and perceived competence were evident. This qualitative phase yielded nuanced insights into how users use generative AI and how trust in AI compares to trust in human responders.

4. Results

The analysis for RQ1 found that all "serviceable" requests identified in the subreddit were directed to the community rather than actual law enforcement or public service providers. These requests, while sometimes involving safety concerns, were managed through the subreddit and did not require immediate police or public service intervention. The types of serviceable requests are summarized in Table 1.

Table 1. Table of serviceable requests on safety related subreddits

Serviceable Requests	Petition support	Job Request	Job search	General guidance request	Child Safety	Risk Reassurance and Anxiety Driven Safety	Regulatory Guidance Seeking	Escalation and Reporting Concerns	Procedural Emergency Guidance
Number of Comments	2	14	15	28	15	9	59	3	19
Percentage	1.2	8.5	9.1	17.1	9.1	5.5	36	1.8	11.6

4.1 Serviceable Requests:

4.1.1 Petition Support

One common request involved petition support, where users asked others to sign and share petitions, often related to safety concerns or causes they believed in. For example, a user shared: "Hi there… petition for technology for children's safety. Please sign and

share (even if you aren't Canadian, this technology is great and will really help).” This type of request calls for community involvement and awareness rather than direct law enforcement or public service action.

4.1.2 Job Requests

Users often sought recommendations for companies or services related to safety. For example, one user requested: “I’m looking for a company that makes short safety videos, 5 mins or below.” Another user asked for help in finding a professional company to create safety forms, explaining: “I work with an electrical contractor and I would like to be able to send my workers to any site and have them fill out specific forms for their safety with little to no prep from me.” These requests pertain to professional services and do not require law enforcement involvement, but instead rely on community knowledge and recommendations in the field of general safety.

4.1.3 Safety Job Search

In the subreddit, many users were actively seeking opportunities and resources related to safety jobs, and the types of requests observed typically involved career development, job searching, and educational resources. These requests did not involve situations that would necessitate police or public service intervention, as they were primarily related to professional growth, job hunting, or learning more about the safety field. Listed below are specific types of safety related job search requests that were identified: Several users sought guidance on job opportunities within the safety field. One user, for instance, asked: “Anyone know a good subreddit for companies looking to hire? I’m leaving the Military with 8 years Health and Safety experience, only an associate’s degree in the field.” This type of request highlights the user’s need for job leads, specifically within the health and safety industry. Similarly, another user inquired: “Thoughts on job market? Anyone have thoughts regarding the job market for safety professionals?” These types of requests reflect a broader concern about the job market within the safety sector and demonstrate how individuals turn to online communities for advice on where to look for job openings and how to navigate the industry. These queries are centered around professional career advice rather than any urgent safety or law enforcement issue.

4.1.4 Guidance Requests

Many requests involved seeking guidance on specific safety procedures or challenges. One user sought advice on dealing with a manager who pressured them to work with asbestos: “What can I do about my manager pressuring me and telling me off for refusing to work with asbestos, even though it's not my job?” This request involved navigating workplace safety and asserting personal rights, though it did not call for police intervention but rather guidance from safety professionals or HR.

4.1.5 Child Safety

Posts related to child safety primarily focused on preventive measures in domestic or recreational settings. These requests were often framed by caregivers who expressed concern about potential hazards rather than active incidents. For example, one user

described staying in a vacation home with young children and highlighted the absence of protective infrastructure: “Hi, I have little kids 3 and 5 and am staying in a house with a pool during an upcoming vacation. There is no fence around the pool so my anxiety is sky high.” The same post continued by seeking advice on specific safety devices: “I started researching pool alarms. Has anyone here had experience with the safety turtle wrist bands?” These posts reflect a proactive orientation toward risk prevention and emphasize the evaluation of safety tools rather than requests for immediate intervention.

4.1.6 Risk Reassurance and Anxiety Driven Safety

A related but distinct category involved risk reassurance and anxiety driven safety requests. These posts were characterized by uncertainty and concern about potential harm, with users seeking validation or probabilistic assessments of risk. For instance, one user asked: “What are the probabilities that I'll get sick from eating an uncooked pizza that was left out overnight, refrigerated, then cooked?” Similarly, anxiety was explicitly articulated in posts involving family safety, as seen in the earlier pool related example where the user stated that their “anxiety is sky high.” These requests were not reports of confirmed danger but rather attempts to reduce uncertainty through reassurance or experiential knowledge from others.

4.1.7 Regulatory Guidance Seeking

Regulatory guidance seeking emerged primarily in professional or occupational contexts, where users requested clarification about safety standards, compliance requirements, or official documentation. One post explicitly referenced regulatory bodies while noting the lack of accessible resources: “Requesting help to find a Construction Site Safety inspection checklist. Title says it all. Really not finding much from OSHA nor MSHA available for download.” The user further questioned whether formal documentation was mandatory: “Is this something that must be created in house, or is there a template floating around?”

4.1.8 Escalation and Reporting Concerns

Another category involved escalation and reporting concerns, where users identified safety hazards but were unsure how to proceed when responsible authorities appeared unresponsive. For example, one user described observing a potential hazard in a public space and expressed uncertainty about next steps: “Does anyone else see a Safety Hazard here? Park staff not concerned. What should I do?” Rather than requesting direct intervention, such posts focused on decision making around escalation, indicating uncertainty about appropriate reporting channels or thresholds for involving authorities.

4.1.9 Procedural Emergency Guidance

Finally, procedural emergency guidance requests centered on hypothetical but plausible emergency situations, emphasizing preparedness rather than realtime response. These posts asked how individuals should act or communicate during emergencies. One example involved interaction with emergency services: “How would I talk to 911 if someone were actively chasing me around my house?”

The results of the content analysis reveal that majority of the serviceable requests are minor and don't require immediate intervention. Therefore, with the advent and proliferation of AI, it is suggested that specialized AI bots would be able to handle such requests with less error and more accuracy than human responses. A major concern is bias and trust in AI bots, therefore, the follow up study looked into this further.

4.2 Trust in AI responses in public safety issues

Qualitative analysis of participant responses revealed that trust in AI responders is conditional, context dependent, and layered, rather than absolute. Participants consistently differentiated between having access to an answer and trusting the answer itself, as well as between low stakes practicality and high consequence decision making. Three higher level themes emerged, each comprising multiple lower level categories.

4.2.1 Conditional Trust Based on Severity and Consequence

Participants' trust in AI varied strongly according to the perceived seriousness and potential consequences of the request.

4.2.1.1 AI as Acceptable for Low Stakes or Everyday Emergencies

For routine or situational needs, participants expressed comfort relying on AI, particularly when immediacy was valued over authority. Several participants redefined "emergency" in practical terms, such as needing quick help for everyday problems. One participant explained, "An emergency is answering my kids' homework on the spot." Similarly, another participant described using AI when dealing with a household concern while alone: "I asked it about an appliance that was making sounds and I was alone at home, at night. It helped." Parents also reported using AI in non critical health contexts involving their children, as one participant stated, "I use it when my kids are sick or when they experience a new symptom" These responses suggest that AI is trusted for immediate reassurance, basic guidance, and everyday problem solving, where the perceived risk is limited.

4.2.1.2. Rejection of AI for High Risk or Medical Emergencies

In contrast, participants expressed clear reluctance or outright refusal to trust AI in situations involving serious health or safety risks. Multiple participants explicitly rejected the use of AI in emergencies, stating, "Never in an emergency," and "No, I don't use it for serious things." One participant demonstrated this skepticism in real time by querying AI during the interview and concluding that "it's not a good source for emergency responses." Concerns were particularly pronounced in medical contexts. A medical doctor participant who used AI for health related information emphasized caution, reflecting a belief that AI lacks the authority and accountability of human professionals when consequences are severe.

4.2.2 Distinction Between Availability of Answers and Reliability of Answers

A central theme across responses was a clear distinction between trust in AI's ability to generate responses and trust in the correctness or appropriateness of those responses. While these might positively correlate in a human, they are not perceived that way in AI bots.

4.2.2.1 Trust in AI's Responsiveness and Knowledge Capacity

Some participants expressed confidence in AI's intellectual and informational capabilities, often positioning it as more knowledgeable than themselves, specially that it is connected to the Internet and constantly updated with new information. One participant stated directly, "Yes, because it's smarter than me." Others emphasized trust in the system rather than the content, particularly when using paid versions. As one participant explained, "I trust the algorithm because I paid for the upgrade version. I know it uses new and updated sources." These responses indicate trust rooted in perceived technical competence, scale of knowledge, and system design, rather than relational trust.

4.2.2.2. Skepticism Toward the Accuracy and Authority of AI Outputs

Despite acknowledging AI's capacity to provide answers, many participants expressed hesitation about relying on those answers for decision making. One participant articulated this distinction succinctly: "I trust it will have an answer, but I don't trust the answer." This sentiment was echoed by others who emphasized that AI responses should not be treated as definitive, especially in contexts requiring expertise or accountability.
This pattern suggests that AI is trusted as an informational resource, but not as an authoritative decision maker, particularly when compared to human responders.
Theme 3: Situational Urgency and Access Driven Trust
Trust in AI increased when human responders were unavailable or too slow, highlighting AI's role as a provisional or stopgap solution.

4.2.3 AI as an Immediate Resource in Time Sensitive Situations

Participants reported turning to AI when speed and accessibility were critical. One participant with a chronic health condition described relying on AI in urgent moments: "I only use it when it's an emergency and I need a response immediately, like a medical issue, because I have an immune system problem and I experience many new situations that sometimes puzzle doctors, so I need answers immediately." In such cases, trust was driven not by full confidence in AI's authority, but by the absence of immediate human alternatives. While AI was valued for accessibility, its limitations also undermined trust. One participant criticized AI responsiveness, noting, "No, because it's slow. It takes its time to process and respond." Another described losing access after heavy reliance: "It helped, but I used it too much and lost access to it." These experiences indicate that trust in AI is also shaped by system performance, availability, and continuity of access, not solely by informational quality.

5. Discussion

The serviceable requests on social media have implications on public services. The content analysis of social media posts (RQ1) revealed several categories of "serviceable" requests that arguably merit attention from police or public service providers. Notably, users frequently sought petition support, made job requests or job search inquiries, and posted general guidance requests. These findings highlight how the general public leverages online networks to address needs that have social or civic dimensions [16],[17],[18]. For instance, calls for petition support suggest that individuals are using social media to rally collective action for a cause a practice aligned with the rise of e petitioning and online advocacy [19]. Through shares and hashtags, a single user's plea for signatures can attract wide visibility, effectively turning a personal request into a community campaign [7]. Public agencies should recognize that such posts may signal grassroots movements or policy grievances that deserve an official response or at least monitoring. Meanwhile, job related requests, such as pleas for employment opportunities or advice on finding work, point to economic and social support needs present in the community. These personal appeals for help with jobs might reflect gaps in existing social services or labor programs [20], prompting questions about whether public workforce agencies could intervene by providing resources or guidance to those turning to social media out of desperation. Finally, the prevalence of general guidance requests where users ask for advice on what to do in challenging situations indicates that people are often seeking expertise or reassurance online for problems that might otherwise fall under the remit of public institutions such as legal advice, health guidance, or safety concerns. In some cases, these guidance seeking posts may be early warning signs of issues; for example, someone unsure how to handle a potentially dangerous domestic situation might seek peer advice before involving authorities. From a public service perspective, these types of posts could be opportunities for engagement or "digital triage." Law enforcement or social service providers, within ethical and legal bounds, might monitor community social media groups to identify such serviceable requests and respond or guide users towards official help channels. This, however, must be balanced with privacy considerations and the need for public trust citizens did not necessarily post to the authorities, so any intervention should be handled with care to avoid perceptions of surveillance.

The interview findings for RQ2 reveal a complex stance toward AI in the role of a responder. Participants generally saw AI as acceptable for low stakes or everyday "emergencies," yet they overwhelmingly rejected AI for high risk or medical emergencies. This dichotomy echoes wider public sentiment, that people are open to AI handling simple, informational, or routine tasks, but they want a human in the loop for serious, high consequence situations [12]. The reasoning expressed by our participants sheds light on why this is the case. Many interviewees acknowledged AI's responsiveness and broad knowledge capacity they trust that an AI system can quickly provide answers and draw upon vast data, which is especially valued in time sensitive situations. Indeed, from a practical standpoint, an AI chatbot is available immediately at any hour, whereas contacting a human official might involve delays or limited hours of operation; participants noted this availability as a major advantage for AI in urgent but not life threatening scenarios. This finding aligns with the idea that AI can enhance accessibility of services by being "always on" [9]. Some even described turning to AI (or imagining doing so) as a first step in an everyday crisis for example, using an AI assistant to get quick instructions on how to shut off a leaking water valve at night

whereas they would reserve calling a human professional for more critical needs. However, alongside this appreciation for availability, there was marked skepticism toward the accuracy and authority of AI outputs. Participants often doubted whether an AI's answer was reliable or fully correct, especially if the issue was complex or dangerous. This skepticism is well founded in light of known AI limitations: a large portion of the public is concerned about AI generated misinformation [11], and our interviewees similarly questioned if an AI could be trusted to give safe advice in, say, a medical emergency. Moreover, even when AI could produce an answer, participants wondered if it had the authority or expertise to be acted upon for example, an AI might instruct someone on a legal matter, but without the credential or accountability of a licensed professional, users hesitated to trust that guidance. These perceptions reaffirm that trust in AI is not only about what the AI says, but who (or what) is delivering the advice; many still instinctively defer to human experts for validation in high stakes contexts.

This study reinforces prior findings in the public safety and social media literature, demonstrating that individuals frequently use online platforms to express concerns and seek advice from peers, advice which they often seriously consider [21]. Despite this, there remains substantial opportunity for further research, particularly regarding the provision of actionable and timely support in online environments. AI driven chatbots may offer a viable mechanism for delivering such serviceable interventions. Moreover, positioning chatbots as social agents through deliberate humanization strategies may enhance their perceived credibility and effectiveness; however, this assumption requires rigorous empirical validation through future research.

5.1 Limitations

This study is subject to several limitations. First, the analysis relied on a limited set of social media posts, which may not represent the full range of safety related requests or platforms used by the public. Second, interview findings reflect self-reported perceptions and hypothetical judgments rather than observed behavior in real emergencies. Third, the study focused on the general public and did not include perspectives from public service professionals or AI system designers. Finally, cultural and contextual factors may limit the generalizability of the findings to other regions or public service settings.

5.2 Conclusion

This study examined whether social media contains serviceable requests that warrant the attention of police or public service providers. The findings show that serviceable requests extend beyond immediate emergencies to include preventive safety concerns, reassurance seeking behavior, regulatory and procedural guidance, and uncertainty around escalation and reporting. The findings further indicate that public trust in AI is highly context dependent. Participants were generally willing to accept AI for low stakes, informational, or preparatory situations that prioritize speed and accessibility. In contrast, trust shifted strongly toward human responders in high risk, medical, or life-threatening scenarios due to concerns about accuracy, authority, accountability, and empathy. The distinction between AI's availability of responses and confidence in their reliability demonstrates a central challenge for AI use in public services. Practically, the results suggest that AI can support public services through information provision and preparedness, provided its role remains clearly bounded and complemented by human

oversight. Expanding AI into domains requiring judgment and emotional sensitivity may risk eroding public trust. Future research should examine institutional perspectives and real-world deployments to better understand how trust in AI assisted public services can be sustained

References

[1] Mahor V, Rawat R, Kumar A, Garg B, Pachlasiya K. IoT and artificial intelligence techniques for public safety and security. InSmart urban computing applications 2023 Feb 7 (pp. 111-126). River Publishers.

[2] Ardabili BR, Pazho AD, Noghre GA, Neff C, Bhaskararayuni SD, Ravindran A, Reid S, Tabkhi H. Understanding policy and technical aspects of ai-enabled smart video surveillance to address public safety. Computational Urban Science. 2023 May 18;3(1):21.

[3] Flahavan E. ChatGOV: Will people trust AI tools to help them use public services? [Internet]. Behavioural Insights Team; 2024 Jan 25 [cited 2026 Jan 30].

[4] Yang Y, Lei P, Pei X, Zhang R, Xiang A. The impact of help-seeking text signals on social support receipt in social media during the COVID-19 outbreak: a cross-sectional study based on signaling theory. Humanities and Social Sciences Communications. 2025 Nov 7;12(1):1-9.

[5] Afroogh S, Akbari A, Malone E, Kargar M, Alambeigi H. Trust in AI: progress, challenges, and future directions. Humanities and Social Sciences Communications. 2024 Nov 18;11(1):1-30.

[6] Almoqbel MY, Alraihan MK, Wohn DY. Understanding decisions to share minor public safety incidents on twitter through a collective action theory lens. International Journal of Social Media and Online Communities (IJSMOC). 2021 Jan 1;13(1):76-92.

[7] Harrison, Teresa M., Catherine Dumas, Nic DePaula, Tim Fake, Will May, Akanksha Atrey, Jooyeon Lee, Lokesh Rishi, and S.S. Ravi. "Exploring E-petitioning and Media: The Case of #BringBackOurGirls." Government Information Quarterly 39, no. 1 (2021): 101569. Accessed January 30, 2026. https://doi.org/10.1016/j.giq.2021.101569.

[8] Sachdeva N, Kumaraguru P. Call for service: Characterizing and modeling police response to serviceable requests on Facebook. InProceedings of the 2017 ACM conference on computer supported cooperative work and social computing 2017 Feb 25 (pp. 336-352).

[9] May K. How local governments can use AI-powered chatbots to improve citizen services [Internet]. Granicus; 2025 [cited 2026 Jan 30].

[10] Aoki N. An experimental study of public trust in AI chatbots in the public sector. Government information quarterly. 2020 Oct 1;37(4):101490.

[11] Siler-Wells S. What Americans Really Think About AI Algorithms: Public Confidence and Transparency in Government [Internet]. Cornell Jeb E. Brooks School of Public Policy; 2025 Nov 12 [cited 2026 Jan 30].

[12] Flanagan A. AI Mental Health Crisis Detection: What 1,000 Americans Really Think About AI-Powered Emergency Response. Iris Telehealth; 2025 Sep 16 [cited 2026 Jan 30].

[13] Tiwari S. Designing Chatbots That Customers Actually Trust [Internet]. CMSWire; 2025 Sep 23 [cited 2026 Jan 30].

[14] Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qualitative health research. 2005 Nov;15(9):1277-88.

[15] Zhao X, Feng GC, Ao SH, Liu PL. Interrater reliability estimators tested against true interrater reliabilities. BMC medical research methodology. 2022 Aug 29;22(1):232.

[16] Purohit H, Castillo C, Imran M, Pandey R. Social-EOC: Serviceability model to rank social media requests for emergency operation centers. In2018 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM) 2018 Aug 28 (pp. 119-126). IEEE.

[17] Chaudhry S, Yuksel M. Engaging communities in public safety via social media. In IEEE INFOCOM 2019-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS) 2019 Apr 29 (pp. 560-565). IEEE.

[18] Almoqbel MY. Exploring, Understanding, Then Designing: Twitter Users' Sharing Behavior for Minor Safety Incidents (Doctoral dissertation, New Jersey Institute of Technology).

[19] Jiang J, Meng T, Zhang Q. From Internet to social safety net: The policy consequences of online participation in China. Governance. 2019 Jul;32(3):531-46.

[20] Almoqbel MY, Wohn DY. Individual and collaborative behaviors of rideshare drivers in protecting their safety. Proceedings of the ACM on Human-Computer Interaction. 2019 Nov 7;3(CSCW):1-21.

[21] Almoqbel M, Wang X, Hiltz SR. Do I care enough? using a prosocial tendencies measure to understand Twitter users sharing behavior for minor public safety incidents.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261099

Evaluating the Effectiveness of AI-Assisted Learning: An Empirical Study Based on Teaching Practice

Ruohan Li, Yongchang Wang, Lu Ma and Tong Xiao [1]
School of Military and Political, National University of Defense Technology, Hunan, China
ORCiD ID: Ruohan Li https://orcid.org/0009-0009-7317-1111

Abstract. This study aims to empirically verify whether AI-assisted learning can improve students' academic outcomes and whether it leads to students' dependency on AI tools. Adopting a quasi-experimental design, this research selected 130 freshmen majoring in Clinical Medicine with similar academic backgrounds to conduct a semester-long comparative experiment across six courses covering both liberal arts and science disciplines. Through a comparative analysis of data on classroom engagement, process assessment, and final exam between the two classes, the study confirmed the hypothesis that AI-assisted learning can effectively enhance learning effectiveness. Furthermore, through an analysis of the experimental group's response latency and AI text similarity of answers during AI-assisted learning activities, as well as their performance in multiple assessments conducted in strictly Non-AI environments, the study concluded that AI-assisted learning did not result in negative cognitive dependency on AI tools among students.

Keywords. AI-assisted learning; Learning Efficacy; Cognitive Dependency

I. Introduction

The proliferation of generative artificial intelligence has triggered a paradigm shift in how students access information and construct knowledge [1]. We define AI-assisted learning (AAL) as an interactive learning process where students engage in continuous and iterative dialogues with generative AI models in the classroom to support their learning and understanding of the teaching content. As AI tools become increasingly integrated into daily study routines, a pervasive consensus prevails that AI-assisted learning—where students actively leverage these tools—offer inherent advantages over traditional, restricted learning settings [2]. However, this prevailing optimism often rests on an unverified implicit premise: that the availability of AI tools invariably translates to enhanced cognitive engagement and deeper knowledge retention [3]. To date, there is a lack of rigorous empirical evidence derived from controlled classroom experiments that isolate the variable of "tool accessibility" to substantiate this assumption. Moreover, alongside the enthusiasm for efficiency, significant concerns remain regarding potential cognitive trade-offs, specifically whether easy access to AI fosters a detrimental

[1] Corresponding Author: Tong Xiao, E-mail: 1964284367@qq.com.

dependency, thereby diminishing students' capacity for independent thinking when the digital "crutch" is removed [4].

Unlike most scholars who study AI empowerment in higher education from the perspective of teachers using AI-assisted teaching [5], this study shifts its focus to the direction of students using AI-assisted learning. Through a one semester teaching practice covering six subjects in arts and sciences, it explores whether AI-assisted learning will have a positive impact on students' learning efficiency and whether it will have a negative impact on cognitive dependence.

2. Research method and subjects

This study employs a quantitative research method utilizing a quasi-experimental design. The sample consists of 130 freshmen from the Clinical Medicine major (Grade 2025) at the National University of Defense Technology, selected for their parallel academic backgrounds. Participants were assigned to two distinct cohorts: Class A (Experimental Group, n=65) and Class B (Control Group, n=65). The study is structured as a one-semester longitudinal teaching experiment conducted during the first academic semester. To ensure the baseline equivalence of the groups prior to the intervention, an independent samples t-test was conducted on the college entrance examination scores of all students. The results indicated no statistically significant difference between Class A and Class B ($t=0.31$, $p=0.76 > 0.05$), confirming that the two groups were highly comparable in academic proficiency. The study is structured as a one-semester longitudinal teaching experiment conducted during the first academic semester.

Among the eight courses offered to university freshmen, two military training courses were excluded as AI tools could not be integrated into their classroom instruction. Consequently, the remaining six courses were selected for the experiment: three liberal arts courses (Fundamentals of Ethics, Traditional Chinese Culture, and Introduction to Psychology) and three science courses (Linear Algebra, Computer Science, and Mathematical Analysis). To minimize the influence of confounding factors on the experimental results, we ensured that both classes utilized identical instructors, teaching content, and pedagogical methods, and that classroom assignments, quizzes, and exam content were completely consistent across the two groups. The sole significant variable distinguishing the experimental group from the control group was that students in Class A were permitted to use AI-assisted tools during learning.

The specific tool used in this experiment was AIZSK, an AI-assisted learning tool developed based on the DeepSeek core architecture. The AI-assisted tool utilized in this study, AIZSK, is a proprietary web-based application developed based on the DeepSeek core architecture. Unlike generic Large Language Models (LLMs) accessed via public interfaces (e.g., ChatGPT or standard DeepSeek chat), AIZSK distinguishes itself in two critical aspects. First, it is deeply integrated with the academy's educational administration system, which facilitates seamless access to and analysis of backend student interaction data. Second, it supports deep learning capabilities based on customized learning resources 'fed' by instructors, such as course syllabi, textbooks, supplementary materials, examination questions, and related academic literature. This integration allows AIZSK to provide context-aware responses tailored to specific curriculum requirements.

In contrast, students in Class B were prohibited from using AI tools to assist their learning in the classroom—a restriction made feasible by the strict military-style

management characteristic of our military academies. Due to the military academy's strict regulations, students are granted access to personal electronic devices (phones/laptops) only on weekends. Crucially, all homework assignments and self-study tasks relevant to this experiment were required to be submitted by 6 p.m. on Friday evening. This structural time lag ensures that the experimental group's in-class AI access was the defining factor affecting their performance, with negligible interference from unsupervised weekend usage.

To ensure that differences in proficiency did not bias the results, the instructor provided standardized training on how to utilize AIZSK for learning prior to the course. Most importantly, to eliminate the variance caused by the adaptation phase or lack of familiarity with the tool, all data used in this study were collected from the one-month period following the completion of this training. By excluding the initial phase, we ensured that the analyzed data reflected a stabilized stage where students were proficient in using the AI tools, thereby minimizing the impact of learning curves on the accuracy of the findings.

3. Can AI-assisted learning improve learning efficiency?

The primary objective of this section is to evaluate whether AI-assisted learning can effectively enhance learning outcomes. To achieve a comprehensive assessment, we analyze the data from three perspectives: classroom engagement, formative assessment, and summative assessment. This multi-dimensional approach allows us to trace the trajectory of student development from behavioral engagement to final academic achievement.

3.1. Comparison of classroom participation between experimental group and control group

To rigorously evaluate the impact of AI-assisted learning on student engagement, this study constructed a multi-dimensional observation framework encompassing both interactive behavior and cognitive status. Specifically, we quantified classroom participation through 3 key indicators: Teacher-Student Interaction, measured by the frequency of discourse; Responsiveness in Learning, characterized by the speed and accuracy of student feedback face-to-face; and Focus Status, assessed via both attendance and sustained attention duration. These metrics were selected not merely to record behavioral frequency, but because they serve as critical proxies for the cognitive activation that AI tools are designed to facilitate during the learning process. Analyzing these dimensions allows us to determine whether students' use of AI tools successfully transitions them from superficial passivity to a state of deep, active engagement.

Table 1. Comparison of student participation in classroom activities in class A and class B

Indicator	Main Dimension	Class A	Class B	t	p	Cohen's d
Teacher-Student Interaction	Average Interaction Times	8.42	3.15	5.68	0.001	0.89
Responsiveness in Learning	Instant Response Rate (%)	96.2	85.5	2.94	0.02	0.64
	Accuracy Rate (%)	82.4	71.8	3.21	0.03	0.72
Focus Status	Attendance Rate (%)	98.5	92.3	2.15	0.05	0.48
	Sustained Attention Duration (%)	91.6	78.2	4.87	0.001	0.82

Based on the data presented in Table 1, a significant divergence in classroom behavior is observed between the two classes, highlighting the effectiveness of AI-assisted learning in stimulating student engagement. First, in the dimension of Teacher-Student Interaction, Class A demonstrated an average of 8.42 interactions per session, which is nearly 2.7 times higher than the 3.15 recorded in Class B. The large effect size (Cohen's d=0.89) suggests that access to AI tools significantly lowered the barriers to participation, encouraging students to shift from passive listeners to active inquirers who utilize technology to scaffold their understanding.

Second, the Responsiveness in Learning metrics reveal a dual advantage for the experimental group. While Class A achieved a remarkably high Instant Response Rate of 96.2%, their Accuracy Rate also reached 82.4%, both significantly outperforming Class B ($p<0.05$). This combination is critical: it indicates that the increased participation was not merely superficial noise, but was driven by enhanced cognitive processing enabled by their learning strategies with AI support. Students in the AI-assisted environment were not only faster to react but also more precise in their grasp of course materials.

Third, regarding Focus Status, Class A exhibited exceptional Sustained Attention Duration at 91.6%, significantly surpassing Class B (78.2%, $p<0.001$). This metric confirms that the AI-mediated learning environment successfully mitigated common distraction issues, fostering a state of deep cognitive immersion where student agency was maintained throughout the session.

In summary, these findings robustly support the hypothesis that AI-assisted learning significantly enhances learning initiative [6]. The experimental group demonstrated superior interaction frequency, higher cognitive responsiveness, and stronger sustained attention compared to the control group in traditional settings. This positive transformation in self-regulated learning behavior lays a necessary behavioral foundation for academic performance. It naturally leads to the subsequent inquiry: whether this heightened engagement and immersion effectively translates into improved mastery of knowledge. To answer this, the following sections will analyze the students' performance in formative assessments and final examinations.

3.2 Comparison of process assessment scores between experimental group and control group

While observable engagement metrics provide evidence of altered classroom dynamics, the true test of any learning intervention lies in its impact on learning outcomes. To evaluate the initial academic efficacy of students' AI-assisted learning, we compared the formative assessment scores between Class A and Class B. Formative assessments, which include quizzes, assignments, and projects, are designed to measure students' ongoing comprehension and skill application throughout the learning process [7]. This analysis seeks to determine whether the heightened engagement observed in Class A directly and immediately translated into superior academic performance during the instructional phase, before examining its long-term effects on knowledge consolidation.

Table 2. Comparison of average scores in students' process assessment in class A and class B

Discipline	Course	M (Class A)	M (Class B)	SD (Pooled)	t	p	Cohen's d
Liberal Arts	Moral Foundations	88	87.2	6.5	0.49	0.63	0.12
	Traditional Chinese Culture	85	86	7.2	1.56	0.12	0.42
	Introduction to Psychology	84.7	85.8	5.8	2.02	0.048	0.52
Science	Linear Algebra	86.2	83.5	8.5	1.32	0.02	0.32
	University Computer Science	84.5	81.2	6.0	1.76	0.04	0.42
	Mathematical Analysis	88	84.1	7.5	2.12	0.038	0.52

Based on the data in Table 2, the impact of utilizing AI tools on students' academic performance during the instructional phase reveals a distinct disciplinary divergence. The experimental group (Class A), who actively used AI support, generally outperformed the control group (Class B) in Science courses, whereas the results in Liberal Arts courses were more mixed.

In the Science category, all three courses demonstrated statistically significant improvements. Mathematical Analysis yielded a highly significant result ($t=2.12$, $p=0.038$), with Class A outscoring Class B by 3.9 points (88.0 vs. 84.1). Similarly, Linear Algebra ($t=1.32$, $p=0.02$) and University Computer Science ($t=1.76$, $p=0.04$) both showed significant differences, indicating that students leveraging AI received substantial support in logic reasoning, calculation, and technical application. The consistent positive outcomes across the Science curriculum suggest that the learning strategy effectively enhanced students' ability to handle complex problem-solving tasks.

In contrast, the impact on Liberal Arts courses was less uniform. Moral Foundations showed a negligible score gap (88 vs. 87.2, $p=0.63$), suggesting that AI-assisted learning provided little incremental value for this subject. Traditional Chinese Culture displayed a non-significant trend favoring the control group ($p=0.12$). Notably, Introduction to Psychology presented a counterintuitive result where Class A (84.7) scored significantly lower than Class B (85.8) ($t=2.02$, $p=0.048$). This negative anomaly suggests that for certain humanities courses, students integrating AI tools into their study workflow may have introduced cognitive interference or an adaptation burden that hindered initial performance. We attribute this negative anomaly to cognitive overload resulting from a knowledge gap between novice learners and advanced AI content. As a discipline with strong scientific attributes, Psychology requires rigorous theoretical scaffolding. However, the AI tools frequently provided comprehensive, high-level theoretical explanations that far exceeded the foundational knowledge of first-year students. Consequently, instead of facilitating understanding, the exposure to overly complex information likely hindered the systematic internalization of concepts in the experimental group, whereas the control group benefited from a learning progression strictly aligned with their novice level.

Collectively, these findings indicate that AI-assisted learning currently offers significant advantages in Science and Engineering disciplines but faces challenges or variable efficacy in Liberal Arts contexts. The inconsistency—ranging from strong gains in Mathematics to a setback in Psychology—implies that the benefits of students' AI adoption are discipline-dependent [8]. This leads to the question of whether these interim disparities reflect the ultimate learning outcomes or if they are transient effects of the adaptation process. To answer this, we must analyze the final examination results presented in the following section.

3.3 Comparison of final assessment scores between the experimental group and the control group

While process assessments provide valuable insights into short-term engagement and incremental skill acquisition, final examinations offer a more robust measure of the ultimate educational outcome: the depth and stability of knowledge retention [9]. This section compares the summative assessment scores of the two groups to determine whether the preliminary gains observed during the instructional phase have solidified into profound academic mastery. The final exam is designed to test not only rote memorization but also the ability to synthesize information and solve complex problems—key indicators of deep learning [10]. If AI-assisted learning truly facilitates the internalization of knowledge structures, we anticipate a significant divergence in performance here, surpassing the subtle variations noted in the process assessments. This comparison seeks to validate the hypothesis that the adaptation cost initially observed in early stages is outweighed by long-term cognitive benefits, leading to superior academic achievements in the experimental group after a full semester of engaging with AI tools.

Table 3. Comparison of average scores in students' final exam in class A and class B

Discipline	Course	M (Class A)	M (Class B)	SD (Pooled)	t	p	Cohen's d
Liberal Arts	Moral Foundations	85.4	78.2	9.0	3.45	0.02	0.82
	Traditional Chinese Culture	83.1	76.5	8.8	3.12	0.03	0.78
	Introduction to Psychology	84.8	77.8	8.5	3.38	0.02	0.80
Science	Linear Algebra	86.2	73.5	11.5	5.21	0.001	1.35
	University Computer Science	84.5	72.8	9.8	4.89	0.001	1.22
	Mathematical Analysis	83.9	74.1	10.2	4.12	0.001	1.05

The data presented in Table 3 reveals a dramatic shift in academic performance compared to the process assessment phase, demonstrating a robust advantage for students utilizing AI support across all disciplines. In the Liberal Arts category, all three courses showed statistically significant improvements ($p<0.05$). Notably, for Moral Foundations, where the process assessment showed negligible difference, the final exam saw Class A improve to 85.4 against Class B's 78.2, with a Cohen's d value of 0.82. This large effect size indicates that the AI-assisted learning approach significantly boosted knowledge retention and conceptual clarity in this subject by the end of the semester.

More striking results were observed in the Science category. All science courses achieved highly significant differences ($p<0.001$), indicating an extremely low probability that these results were due to chance. In Linear Algebra, Class A outperformed Class B by a substantial margin of 12.7 points (86.2 vs 73.5). The corresponding Cohen's d value of 1.35 is particularly noteworthy; as an effect size exceeding 1.0, it suggests a transformative impact of students' use of AI tools on their ability to master complex logic and calculation. Similar large effect sizes were recorded for University Computer Science ($d=1.22$) and Mathematical Analysis ($d=1.05$), confirming that the benefits in science-oriented courses were not merely marginal but substantial.

In contrast to the incremental or inconsistent improvements noted during the process assessment, the final examination results provide compelling evidence that AI-assisted learning effectively enhances students' overall academic proficiency. However,

a closer examination of the data uncovers a critical disciplinary disparity: the impact of students' AI adoption appears significantly more potent in Science disciplines than in Liberal Arts. While the Liberal Arts courses demonstrated solid improvements with effect sizes ranging around 0.8 to 0.82, the Science courses exhibited substantially larger effect sizes, with all exceeding 1.0 and peaking at 1.35 in Linear Algebra. This pronounced difference suggests that AI tools may be uniquely effective for students in facilitating the comprehension of abstract logical structures, complex algorithms, and mathematical modeling—cognitive tasks that are heavily emphasized in science curricula—compared to the more interpretative and discourse-based nature of liberal arts subjects [11]. Consequently, while the study validates the general efficacy of AI-assisted learning, it also highlights that the degree of enhancement is subject-specific, with Science disciplines deriving a more profound advantage from the integration of AI technology.

4. Will AI-assisted learning lead to students' dependence on AI tools?

AI-assisted learning is frequently accompanied by concerns regarding the crutch effect—the risk that excessive reliance on AI might erode students' cognitive autonomy and foster a problematic dependency [12]. To rigorously investigate this issue, it is insufficient to rely solely on self-reported surveys, as students often lack the metacognitive awareness to accurately gauge their own reliance. Instead, this section targets the objective reality of their learning behaviors through a two-dimensional empirical analysis. By cross-referencing shifts in behavioral patterns with actual cognitive proficiency under unassisted conditions, we can robustly determine whether AI integration in their study routines fosters constructive tool usage or leads to a counterproductive loss of independence.

4.1 The reaction time of experimental group students using AI tools and the AI similarity of their answers

To objectively assess the extent of students' engagement with AI tools and the presence of behavioral dependency throughout the semester, we extracted learning data from the AIZSK backend at the end of each month. We quantified two metrics: the average response latency—defined as the time interval between the presentation of a learning prompt and the student's final answer submission—and the AI similarity of their submissions—measured using the AIGC detection tool provided by the VIP paper check system. These two indicators were utilized to examine whether students utilized AI as a learning scaffold (i.e., engaging in deep cognitive processing before using AI as a supplement) or exhibited over-reliance on AI (characterized by abnormally fast submission speeds and a lack of originality). We posit that latency time serves as an effective proxy for cognitive effort when analyzed with text similarity. A "shallow" interaction is characterized by a specific signature: minimal response time coupled with high text similarity to the AI output, suggesting superficial copying without critical scrutiny. Conversely, "deep" engagement is inferred from increased latency, as independent thinking requires time to interrogate, synthesize, and reconstruct content. Thus, tracking the dynamic correlation between these variables allows us to distinguish between passive reliance and active cognitive processing, ensuring prolonged latency reflects thought rather than indecision. However, we acknowledge limitations: increased

latency may occasionally stem from external distractions or students' unnecessary procrastination. Nevertheless, aggregate trend data across the study period robustly reflect general engagement patterns, minimizing the impact of such anomalies. Following the five-month instructional period, the collected data are organized as presented in Table 4.

Table 4. Student interaction metrics (response latency and ai similarity) across five phases

Discipline	Course	Phase I	Phase II	Phase III	Phase IV	Phase V
Liberal Arts	Moral Foundations	20S/85%	27S/82%	35S/80%	37S/77%	40S/50%
	Traditional Chinese Culture	23S/88%	26S/84%	32S/78%	35S/66%	38S/56%
	Introduction to Psychology	24S/87%	29S/83%	34S/73%	36S/61%	39S/55%
Science	Linear Algebra	20S/90%	30S/85%	33S/81%	35S/69%	41S/65%
	University Computer Science	25S/88.5%	28S/83%	32S/77%	37S/65%	42S/64%
	Mathematical Analysis	28S/86%	29S/80%	34S/78%	38S/67%	40S/62%

The data presented in Table 4 reveals a consistent behavioral evolution across both Liberal Arts and Science disciplines throughout the five-month observation period. We observe a clear inverse relationship between response latency and AI similarity: as the semester progressed, students consistently spent more time formulating their answers, resulting in outputs that became increasingly distinct from the AI's raw generation. For instance, in Introduction to Psychology, response latency nearly doubled from 24s in Phase I to 39s in Phase V, accompanied by a precipitous drop in AI similarity from 87% to 55%. This trend—replicated across courses like Linear Algebra (20s to 41s; 90% to 65%)—indicates a decisive departure from the copy-paste model. The trajectory of the data demonstrates that as students allocated more time to the task, their outputs became increasingly distinct from the AI's raw generation, effectively ruling out the hypothesis of mindless or uncritical dependency.

However, it is essential to acknowledge the ambiguity of these metrics. While the data refutes the model of uncritical rote copying—as the declining similarity rates preclude the possibility of a purely copy-paste approach—it cannot definitively prove deep learning. The increased latency and textual modification may indicate genuine cognitive engagement, but conversely, they may reflect a surface-level compliance strategy where students merely rephrase AI outputs to meet evaluation criteria (e.g., avoiding plagiarism detection) without fully internalizing the underlying knowledge. The metrics confirm a shift in behavior, but without further qualitative evidence, we cannot definitively distinguish between rephrasing for correctness and rephrasing for understanding.

In conclusion, the analysis of Table 4 suggests that the adoption of AI-assisted learning does not inevitably lead to a static state of simple reliance. Within the scope of the current quantitative data, there is evidence that students adapted their interaction patterns over the semester, moving away from direct replication towards more independent text construction. However, while the rejection of a copy-paste model is statistically robust, we must exercise caution in equating this behavioral shift with deep cognitive mastery. The increased latency and reduced similarity indicate a deviation from passive reliance, yet they do not definitively reveal whether this independence stems from a genuine desire to learn or merely a strategy to comply with evaluation standards. To determine whether this shift represents profound cognitive engagement or superficial

compliance, it is necessary to cross-reference these behavioral patterns with qualitative evidence, which will be examined in the following section.

4.2 Comparison of test scores between experimental group and control group students in a non-AI environment

To corroborate the behavioral patterns observed in the previous section and determine whether interaction with AI tools has a substantive impact on students' intrinsic cognitive capabilities, we conducted a series of unassisted proficiency assessments. Consistent with the monthly data collection points, these five tests were administered at the end of each month in a strictly Non-AI environment, eliminating any external digital support. Crucially, the design of these assessments deliberately avoided objective questions—such as multiple-choice or fill-in-the-blank items—that rely on rote memory or standard answers. Instead, the test content was constructed to specifically evaluate divergent thinking and creative problem-solving abilities, requiring students to construct subjective arguments, design original solutions, and apply knowledge flexibly. By focusing on higher-order cognitive skills that AI tools typically struggle to replicate authentically, this testing methodology ensures that student performance accurately reflects their independent intellectual capacity [13]. This design allows us to rigorously test our hypothesis: if students have become psychologically or cognitively dependent on AI, their performance in these creative, unassisted tasks would likely stagnate or decline compared to the control group.

Table 5. Comparison of average scores between class a and class b in a non-AI testing environment

Discipline	Course	The first test	The second test	The third test	The fourth test	The fifth test
Liberal Arts	Moral Foundations	80/81	81/81.5	83/82.1	83/82.1	84/83.2
	Traditional Chinese Culture	82.9/82.5	83/83.3	84/83.3	83/83.5	84/82.8
	Introduction to Psychology	82.8/83	83/84	83.1/83	85/86.5	84/84.2
Science	Linear Algebra	81.7/81.5	83.2/83	84/83	85.5/84	86/84.2
	University Computer Science	82.5/82.1	83.5/83	85/83.1	86/84	86.6/85
	Mathematical Analysis	84.5/85.2	85/84	85.5/84	86/83.5	88/84

The data presented in Table 5 provides a definitive assessment of students' intrinsic cognitive capabilities in a non-AI environment, revealing a distinct performance gap between the experimental group (utilizing AI tools) and the control group (traditional instruction).

Across all six disciplines, students in the experimental group consistently outperformed their counterparts in the control group, particularly in the latter stages of the semester. In the Liberal Arts category, courses such as Introduction to Psychology show a clear widening of the proficiency gap: while the control group's score increased modestly from 83 to 84.2 over the five tests, the experimental group demonstrated more robust growth, peaking at 85 in the fourth test and finishing consistently higher. Similarly, in the Science disciplines, the disparity is more pronounced. For Mathematical Analysis, the experimental group improved from 84.5 in the first test to 88 by the fifth, whereas the control group actually exhibited a decline from 85.2 to 84. This trend is mirrored in University Computer Science, where the experimental group maintained a steady upward

trajectory (82.5 → 86.6), eventually securing a lead of 1.6 points over the control group by the final assessment.

These results are particularly significant given the nature of the test content. As the assessment was designed to evaluate divergent thinking and creative problem-solving—skills that are difficult to simulate through rote memorization—the superior performance of the experimental group suggests that their interaction with AI tools facilitated a deeper internalization of knowledge. The consistent upward trend in the experimental group's scores, contrasted with the relative stagnation or occasional decline observed in the control group, indicates that the integration of AI tools into the learning process did not lead to intellectual atrophy. Instead, it appears to have fostered a higher level of cognitive engagement, enabling students to independently generate more original and complex solutions than those restricted to traditional learning methods.

The trajectory of test scores serves as a critical counterpoint to the ambiguity noted in the behavioral metrics of Section 4.1. While the rephrasing observed in the AI logs could have theoretically been a superficial compliance strategy, the consistently higher scores in objective, Non-AI creative assessments make this possibility less likely. The data demonstrates that students using AI tools were able to transfer their learning experiences into improved independent performance. Therefore, based on the current evidence, we can cautiously conclude that AI-assisted learning, when implemented as described, does not compromise students' independent thinking abilities. Furthermore, the observed performance advantage suggests that AI-assisted learning may play a constructive role in enhancing proficiency in creative and divergent thinking tasks, although strictly attributing this improvement solely to the intervention requires further investigation to rule out other contributing factors.

5. Conclusion

This study provides an empirical investigation into whether AI-assisted learning can enhance academic proficiency and induce AI dependency within a higher education context. Based on the analysis of behavioral logs and assessment scores, we draw the following key conclusions.

First, our findings indicate that engaging in AI-assisted learning effectively improves overall academic proficiency. Notably, the data reveals a crucial disciplinary variation: the impact of AI intervention was significantly more pronounced in science disciplines compared to liberal arts. We attribute this disparity to the fundamental epistemological differences between disciplinary domains [14]. Science disciplines, characterized by structured problem-solving, algorithmic logic, and definitive solutions, align closely with the operational strengths of current generative AI. In these contexts, AI functions effectively as an advanced scaffolding tool, rapidly resolving technical bottlenecks and allowing students to focus on higher-order conceptual integration. Conversely, liberal arts education often prioritizes critical interpretation, subjective argumentation, and nuanced textual analysis—cognitive processes that require deep contextual understanding and humanistic insight, areas where AI models may offer less precise or relevant augmentation [15]. Thus, the efficacy of AI-assisted learning appears to be contingent upon the structural affinity between the tool's algorithmic capabilities and the cognitive demands of the specific discipline.

Second, regarding the concern of cognitive offloading, the study finds that AI-assisted learning does not compromise students' independent thinking abilities. On the

contrary, the evidence suggests that AI tools may play a constructive role in fostering students' creative and divergent thinking [16]. While we must remain cautious in strictly attributing the magnitude of this improvement solely to the AI intervention—given the potential influence of other covariates—the results robustly confirm that the integration of AI tools into students' learning routines, at the very least, serves as a benign and likely beneficial supplement to traditional learning.

Despite these promising insights, several limitations must be acknowledged. First, the study was conducted over a single semester; thus, it may not fully capture the long-term effects of AI-assisted learning. It remains unclear whether the observed behavioral adaptations will be sustained or if AI dependency might emerge after a more prolonged period of exposure. Future longitudinal studies are needed to monitor these trajectories. Second, as the participants and specific courses involved in this experiment were limited, the generalizability of our conclusions is inherently constrained. Future research should expand the scope to diverse educational settings to validate these findings.

References

[1] Xiumei Lv. Inquiry into the Human-Machine Collaborative Teaching Model of Sora Empowering University Ideological and Political Theory Courses. Theoretical Observation. 2024 Sep; 9:38-41.
[2] Jingyi Hu. Current Status and Challenges of AI-Assisted Teaching in the AI Era. Technology and Innovation. 2021 Feb; 2:149-150.
[3] Liya Tu. Cultivating Creative Thinking: Starting from Reforming the Education System. Shanghai Education. 2025 Oct; 29:30-34.
[4] Chenhui Gao, Liyuan Liu, Jianli Wang. Research on Design Teaching Model Under AI Intelligent Assistance: The Role and Impact of Artificial Intelligence. Art and Design. 2024 Feb; 2(3):125-128.
[5] Dalong Gong, Shu Zhang. Analysis of the Expanding Effect of University Informatization Construction on the Course Quality Evaluation System. China New Telecommunications. 2025 Oct; 27(20):68-70.
[6] Chen Li, Shuwang Wang. Exploration of Intelligent Assisted Teaching in Higher Vocational Colleges Based on AI Large Models. Office Automation. 2025 Nov; 30(11):37-39.
[7] Bin Zhao. Reshaping the Classroom: From Blended Teaching to AI-Assisted Smart Teaching. China University Teaching. 2025 Nov; 11:69-75.
[8] Wei Guo, Zhihua Zhan. Breakthrough and Reconstruction: AI Empowering Ideological and Political Education. Moral Education in China. 2025 Jun; 17:16-21.
[9] Shuai Liu. Research on Agent Teaching Assistants Improving Interdisciplinary Adaptability: Based on Constructivist Learning Theory. Herald of Journalism and Communication. 2025 Aug; 16(22):153-156.
[10] Xiaoqin Shi, Lu Liu, Ni Zhu. Exploration of AI-Assisted Personalized Learning Path Design in Blended Teaching. Internet Weekly. 2025 Sep; 12:34-36.
[11] Yuze Zhao, Yuanxing Ye, Yuefang Zhao, Jun Jiang, Shang Yu. Construction of Smart Teaching System for Ecology Practical Courses Based on AI-Assisted Recognition. China Modern Educational Equipment. 2025 Jun; 23:58-61.
[12] Zhe Yang, Longhai Rao. Online Course Resource Design Based on Cognitive Load Theory. Journal of Electrical & Electronic Education. 2025 Apr; 47(4):138-142.
[13] Aiyun Fu, Bing Liu. On the Cultivation of College Students' Innovation Ability in Smart Learning Environment. Journal of Social Science of Jiamusi University. 2025 Jul; 43(7):145-146.
[14] Shaoquan Wen. Exploration and Practice of AI-Assisted 'Advanced Mathematics' Classroom Teaching. Intelligence. 2025 Sep; 24:1-4.
[15] Weiyu Ye, Jing Chen. Status and Research of AI-Assisted Teaching in the AI Era. Technology Wind. 2025 Apr; 4:68-70.
[16] Hongyu Shi. Research and Application of AI-Assisted Teaching Tools. Journal of Jilin Teachers Institute of Engineering and Technology. 2025 Feb; 41(2):46-53.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)

© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261100

Interactive Genetic Algorithm for Narrative-Driven VR Scene Layout Optimization

Yan Zhang[a,1] ,Wei Chen[a,2*] and Gang Yang[b]
[a]*Department of Arts and Media, Century College, Beijing University of Posts and Telecommunications, Beijing, China*
[b]*School of Information Science & Technology, Beijing Forestry University, Beijing, China*
ORCiD ID: Yan Zhang https://orcid.org/0000-0001-8356-9507

Abstract. The integration of large-scale models and intelligent algorithms into virtual reality (VR) visualization of magical realism literature - including intelligent VR film and television production and AIGC-driven VR gaming-has not only redefined the representational paradigms of literary imagery but also sparked significant scholarly discourse on the interplay between technological empowerment and humanistic aesthetics. This study employs the VR scene generation of *Not in This Lifetime* [1], a science fiction work by the corresponding author, as a research case to develop a comprehensive algorithmic framework that integrates multimodal perception, aesthetic principles, and narrative logic to propose a full-chain framework for VR scene layout generation. This work proposes an interactive genetic algorithm-based optimization method for large-scale 3D building cluster layout as its core innovation. Following the "text-feature-model-layout" pipeline, it analyzes algorithm-driven aesthetic evolution, the tension between technical rationality and artistic expression, and humanistic value retention in immersive scenarios. A multi-objective fitness function integrating visual saliency, aesthetic norms, narrative logic and functional constraints is established. Validated on ten literary scenes, the proposed method exceeds baseline schemes in narrative consistency, aesthetic performance and 3D layout constraint adaptation. It achieves an NFMR of 89.7% (37.58% higher than GA) and a 7.86 aesthetic score, with 30-participant tests confirming its validity in literary immersive visualization. Combined with narrative constraints and dynamic feedback, the system balances computation efficiency and artistic expression, delivering a robust technical paradigm for cross-media adaptation of literature and VR interactive contents.

Keywords. Virtual reality; interactive genetic algorithm; 3D architectural layout optimization; interdisciplinary integration

[1] First Author: Yan Zhang, Associate Professor at the School of Art and Media, Beijing University of Posts and Telecommunications Century College. His research focuses on interdisciplinary aesthetic integration, virtual reality technologies, and 3D game design. email: ryanchang_cuc@yeah.net.

[2] *Corresponding Author: Wei Chen, Vice Dean and Professor at the School of Art and Media, Beijing University of Posts and Telecommunications Century College. Her primary research areas include animation and game design, virtual reality, and aesthetic theory. email: 1773676407@qq.com.

1. Introduction

The immersive capabilities of virtual reality (VR) technology provide a novel medium for expressing science fiction and magical realism literature, enabling the concrete visualization of abstract literary imagery. This study is grounded in the science fiction work *Not In This Lifetime*, published as an outcome of an art foundation project led by the first author, and investigates an intelligent framework for the rapid generation of VR scenes based on literary narratives [1]. The narrative environments in the original work−characterized by regional specificity, cybernetic elements, and magical realist aesthetics−not only constitute a complex and challenging domain for VR implementation but also reflect broader transformations in literary dissemination in the digital era. As algorithms increasingly participate in the decoding, transformation, and algorithmic mediation alters the representational modes of literary imagery, fundamental aspects of literature − including its essential attributes, aesthetic criteria, and dissemination paradigms−are undergoing profound reconfiguration. The proposed framework is not limited to the magical realism work *Not in This Lifetime*. It can be extended to science fiction, fantasy, and regional literature by adjusting the semantic tagging system and aesthetic constraints. The core "text-narrative-layout" pipeline and interactive genetic algorithm are universal for text-driven 3D scene generation tasks. This research systematically performs innovative investigations and explorations to address the aforementioned issues.

2. Background

Recent advances in VR scene generation have improved accuracy and contextual fidelity. Zhang et al. [2] integrated visual saliency and aesthetics into a 3D viewpoint estimation method. Bartolomeo et al. [3] clarified neural attention mechanisms and visual perception foundations; Itti et al. [4] proposed the center-surround mechanism−a core computational aesthetics model for scene optimization. Conventional layout algorithms prioritize functional constraints but overlook narrative structure and aesthetic intent, yielding technically sound yet humanistically shallow outputs. IGAs excel in multi-objective optimization with subjective criteria, with applications in artistic creation [5], architectural form design [6], indoor layout evaluation [7], ecological landscape simulation [8], and mixed-reality scene synthesis [9]. However, they scale poorly to large 3D building clusters and lack deep integration with literary narrative logic [10,11,12]. Using *Not In This Lifetime* as a case study [13], we propose a full-chain VR scene generation framework that unifies multimodal perception, aesthetic rules, and narrative semantics. Under the guidance of aesthetic computing, our IGA embeds narrative-aesthetic constraints to enable precise, context-aware layout optimization−balancing algorithmic rigor with literary expressiveness while reducing narrative alienation and advancing immersive literary visualization.

3. A Semantic and Aesthetic-Driven Algorithmic Framework for Scene Layout

3.1. Text Semantic Decomposition Module

To achieve accurate visualization of literary imagery in *Not In This Lifetime*, this work incorporates narrative logic into the technical framework. The novel portrays cultural landscapes of the North China Plain and socio-spatial transformations projected into a cyber-evolved future, whose evocative expressions pose challenges for computational interpretation (Figure 1). We propose a four-dimensional tagging framework−regional characteristics, cyber elements, magical imagery, and narrative functions−that advances conventional keyword extraction. Using the Tencent Hunyuan 3D large model, the framework converts magical-realist text segments into computable narrative features for 3D scene generation while preserving semantic depth and aesthetic nuance.



Figure 1. The decaying marketplace (left) and the cyberpunk-style temple (right) generated based on the text of the monograph.

3.2. Interactive Genetic Algorithm Implementation Module

This algorithm uses 3D grid visual saliency, grid curvature, and aesthetic principles−the rule of thirds and the golden ratio−as evolutionary genotypes. Guided by an interactive genetic algorithm [14], it optimizes spatial configurations of large-scale 3D building clusters under narrative constraints from *Not In This Lifetime*, ensuring both technical performance and narrative fidelity.

- **Chromosome Coding Model**

We adopt a matrix coding scheme: the chromosome Chrom is an $m \times n$ matrix, where each gene $Gene_{a,\beta}$ encodes a building module's full parameters $Gene_{a,b} = [ID, (x, y, z), \theta_h, \theta_v, S, K_{avg}, C_{rule3}, C_{golden}, B_{total}, F_n]$. Here, (x, y, z) are relative coordinates within the unit block; $\theta_h \in [0°, 360°]$ and $\theta_v \in [0°, 15°]$are horizontal and vertical orientation angles; S and K_{avg} are normalized to [0, 1]; C_{rule3}, C_{golden}, andB_{total} measure adherence to the rule of thirds, golden ratio, and visual balance; F_n is a one-hot narrative function label.

- **Fitness Function Model**

We formulate a multi-objective fitness function that integrates saliency-guided attention, aesthetic quality, narrative adaptation, and functional constraints−computed as a weighted sum:

$$F_{fitness} = \omega_1 F_{salient} + \omega_2 F_{aesthetic} + \omega_3 F_{narrative} + \omega_4 F_{constraint} \quad (1)$$

The significance-guided fitness calculation is presented in Equation (2):

$$F_{salient} = \frac{1}{N_{building}} \sum_{k=1}^{N_{building}} S_k \cdot sim(Loc_k, S_{focus}) \quad (2)$$

Here, S_k denotes the saliency weight of the k-th building; $sim(Loc_k, S_{focus})$ represents the spatial alignment between the building location Loc_k and the saliency focus region S_{focus}; $N_{building}$ refers to the total number of buildings.

The aesthetic quality score is computed as defined in Equation (3):

$$F_{aesthetic} = \eta_1 C_{rule3} + \eta_2 C_{golden} + \eta_3 B_{total} + \eta_4 C_{color} \quad (3)$$

Here, C_{color} denotes the color harmony metric, which is extracted using the improved color pattern bag algorithm and normalized to the interval $[0, 1]$; η_1, η_2, η_3, and η_4 are weighting coefficients with values of 0.3, 0.25, 0.25, and 0.2, respectively.

Narrative fit is computed according to Equation (4):

$$F_{narrative} = \frac{1}{M}\sum_{t=1}^{M} match\,(F_{n,k}, T_t) \quad (4)$$

T_t is the narrative function label for category t; $match(\cdot)$ outputs 1 for a match, 0 otherwise; and M is the total number of narrative function categories. Functional constraint satisfaction is defined in Equation (5).

$$F_{constraint} = \frac{1}{4}(dist(d_k) + ori(\theta_k) + access(P_k) + comp(T_k)) \quad (5)$$

Here, $dist(d_k)$ measures building spacing compliance (1 if $d_k \geq 1.5\times avg_{width}$, else d_k /(1.5× avg_{width})); $ori(\theta_k)$ quantifies orientation compatibility; $access(P_k)$ evaluates transportation accessibility; and $comp(T_k)$ assesses texture–style compatibility. All metrics are normalized to [0, 1]. Weights $\omega_1 = 0.3$, $\omega_2 = 0.3$, $\omega_3 = 0.2$, and $\omega_4 = 0.2$ are derived via the Analytic Hierarchy Process (AHP). The weight coefficients are determined by 5 experts in digital media and 3 experts in architectural design through pairwise comparison. The consistency ratio CR = 0.0078 < 0.1, which meets the AHP consistency test standard, proving the rationality of weight distribution.

- **Genetic Operators**

Selection: Use tournament selection ($k = 5$) to select parents from random samples; repeat until 9 parents are selected, then generate offspring via crossover and mutation. *Crossover*: Block crossover ($p_c = 0.75$) swaps genes within a rectangular region $R = [a_1, a_2] \times [b_1, b_2]$ (where $a_1 < a_2$, $b_1 < b_2$) of the chromosome grid. Offspring must satisfy $C_{golden} \geq 0.5$ (Iterate up to 10 times); otherwise, directly use the parent individual as the offspring. *Mutation*: Compound mutation ($p_m = 0.04$) perturbs a randomly selected set of genes: Coordinates: $x' = x + \Delta x$, $y' = y + \Delta y$, $z' = z + \Delta z$, with Δx, Δy, $\Delta z \in [-0.1 \cdot x_{max}, 0.1 \cdot x_{max}]$; Significance weight $S' = S + \Delta S$, $\Delta S \in [-0.2, 0.2]$, $S' \in [0, 1]$; Rule-of-thirds flag $C_{rule3}{}' = 1 - C_{rule3}$ (binary flip); Orientation $\theta_h{}' = \theta_h + \Delta\theta_h$, $\Delta\theta_h \in [-15°, 15°]$.

3.3. Adaptive Fusion Guided by Aesthetic Principles and Scene Narrativity

To address scale heterogeneity in large-scale architectural ensembles, we introduce a scale-adaptive weight $\omega_{scale} = Size(k)/AvgSize$ that dynamically modulates each genotype feature's contribution (As shown in Equation (6)).

$$F_{gene}(k) = \omega_{scale}(k) \cdot \left(\mu_1 S(k) + \mu_2 K_{avg}(k) + \mu_3 C_{aesthetic}(k)\right) \quad (6)$$

Here, $Size(k)$ is the geometric scale of building k; $AvgSize$ is the ensemble mean scale; $\mu_1 = 0.4$, $\mu_2 = 0.3$, $\mu_3 = 0.3$ are the baseline weights; and this balances genotype feature fusion despite inter-building scale variation. A dual-constraint verification function $ConstraintCheck(\cdot)$ is embedded in crossover and mutation to ensure narrative coherence and aesthetic compliance. Formally, $ConstraintCheck(Gene') = 1$ if and

only if $F_{narrative}(Gene') \geq 0.6$ and $F_{aesthetic}(Gene') \geq 0.5$; else, its value is 0. If $ConstraintCheck(Gene') = 0$, the operator retries—improving constraint satisfaction and evolutionary robustness.

- **Attention-Guided Fitness Weighting Strategy**

This strategy uses saliency predictions to dynamically weight visually salient regions in building clusters, yielding an attention-modulated fitness function:

$$F_{fitness}^{'} = F_{fitness} \cdot \left(1 + J_{region}(k) / \max(J_{region})\right) \quad (7)$$

Here, $J_{region}(k)$ is the average saliency value over the region occupied by building *k*; this strengthens attention-guided layout optimization.

3.4. Dynamic Feedback-Driven Optimization Model

A dynamic adjustment mechanism—grounded in real-time user interaction signals and narrative progression metrics—is formalized in Equation (8):

$$Param(t) = Param(t-1) + \lambda \cdot UserFeedback(t-1) + (1-\lambda) \cdot NarrationProgress(t-1) \quad (8)$$

Here, $Param(\cdot)$ denotes the parameter vector for building location, orientation, and saliency weight. $UserFeedback(t)$ is a vector of real-time interaction signals at step t, computed from synchronized eye-tracking and head-pose data and dimensionally matched to $Param(t)$. $NarrationProgress(t)$ is a vector of narrative-phase weights at step t, where each element is mapped to a fixed phase value: 0.1 (lyrical), 0.2 (dramatic), 0.3 (conflict), and 0.4 (climax). The value $\lambda = 0.6$ is empirically set to balance the contribution of user feedback, based on user study results. When the narrative phase transitions, $NarrationProgress(t)$ updates in a stepwise manner to the corresponding phase value.The MLP architecture for user feedback quantification is: Input layer (user feedback + narrative progress) → 2 hidden layers (128 and 64 neurons, ReLU activation) → Output layer (parameter adjustment vector). Hyperparameters: Batch size = 16, learning rate = 0.001, optimizer = Adam, epochs = 50. Training data: 1,200 sets of user interaction data including eye-tracking and head-pose signals collected in the pre-experiment.

3.5. Algorithmic Workflow

The proposed algorithm comprises three phases: (1) Preprocessing: extracts genotype features (multi-scale visual saliency, convexity/concavity, and aesthetic features); defines layout, narrative, and aesthetic constraints; and generates 10 initial layout candidates. (2) Iterative optimization: calculates fitness based on saliency, aesthetics, narrative coherence, and functional constraints; quantifies user feedback via the multi-layer perceptron (MLP); performs selection, block crossover, and compound mutation (followed by the ConstraintCheck(·) verification with up to 10 retries for invalid offspring); and adapts to user feedback until the relative fitness improvement (against the previous generation) falls below 1% or 50 generations are reached. (3) Output: generates final layouts satisfying aesthetic, narrative, and functional goals, parameterized by position, orientation, scaling, and camera-path specifications.

4. Comparative Experiment Design

4.1. Experimental Setup

Hardware: Intel Core i9-12900K CPU, NVIDIA GeForce RTX 3090 GPU, 64 GB DDR5 RAM, and 2 TB NVMe SSD; Software: Python 3.8.18, PyTorch 1.12.1+cu116, CUDA Toolkit 11.6.2, Unity Editor 2023.2.20f1 (for real-time 3D scene rendering), and OpenCV 4.5.5 (for image preprocessing and feature extraction).

4.2. Algorithm Selection for Comparison

To validate the proposed algorithm, we compare it against four baselines: classic GA (without narrative or aesthetic constraints), PSO for scene layout, a greedy algorithm (guided solely by aesthetic rules), and IGA-Unconstrained (lacking narrative constraints and multi-scale adaptation). The custom dataset contains 3D models of 10 canonical magical realism scenes from *Not In This Lifetime*, each instantiated with 20–50 parametric building modules annotated with structured metadata: narrative function (Figure 2), aesthetic attributes, and humanistic associations. We have released all structured prompt templates, prompt examples, and parameter conversion tables for VR scene generation publicly in Zenodo at DOI: https://doi.org/10.5281/zenodo.19905900. These open materials allow full reproduction of our semantic parsing, parameter mapping, and generation pipeline without relying on copyrighted 3D assets. Narrative labels are jointly annotated by 2 literature researchers and 1 VR designer, with inter-annotator agreement (Cohen's Kappa = 0.85).

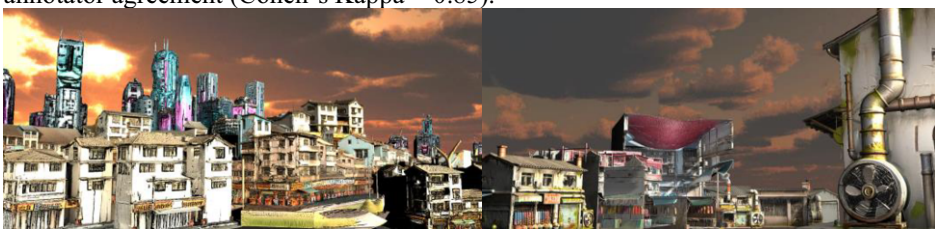


Figure 2. A scene layout excerpt derived from a landscape node in the original text

4.3. Evaluation Metrics

We evaluate performance across three independent dimensions: (1) Aesthetic quality: C_{rule3} (rule-of-thirds), C_{golden} (golden ratio), B_{total} (visual balance), and C_{color} (color harmony), plus mean 10-point Likert scores from 30 domain experts (1 = poor, 10 = excellent). (2) Narrative fit: Narrative function matching rate, theme understanding score (BERT-based), and coherence rating (mean expert score on temporal/logical continuity). (3) Functional constraint satisfaction: Building spacing compliance rate, traffic accessibility (10-point expert rating), and orientation compatibility (10-point expert rating).

4.4. Experimental Results and Analysis

- **Aesthetic Quality Comparison**

We evaluated aesthetic generation across five algorithms using objective metrics and subjective human assessment (Figure 2). Our method achieved C_{rule3} = 0.887, C_{golden} = 0.856, B_{total} = 0.879, and C_{color} = 0.882−outperforming GA and PSO significantly. Against IGA-Unconstrained (C_{rule3} = 0.823,C_{golden} = 0.795,B_{total} = 0.812, C_{color} = 0.836), it improved C_{rule3} by 7.78%, C_{golden} by 7.67%, B_{total} by 8.25%, and C_{color} by 5.50%. In subjective aesthetic scoring (SAS), it scored 7.86−surpassing GA (6.23), PSO (6.58), Greedy Algorithm (7.35), and IGA-Unconstrained (7.62) by 1.63, 1.28, 0.51, and 0.24 points, respectively. These results confirm its superior visual coherence, color harmony, and alignment with human aesthetic preferences (Table 1).

Table 1. Aesthetic Quality Comparison Results (SAS: Subjective Aesthetic Score)

Algorithm	C_{rule3}	C_{golden}	B_{total}	C_{color}	SAS (points)
GA	0.682	0.654	0.693	0.721	6.23
PSO	0.715	0.687	0.725	0.743	6.58
Greedy Algorithm	0.786	0.762	0.778	0.795	7.35
IGA-Unconstrained	0.823	0.795	0.812	0.836	7.62
Proposed Algorithm	0.887	0.856	0.879	0.882	7.86

- **Narrative Adaptation Evaluation**

We evaluated narrative adaptation using three metrics: Narrative Function Matching Rate (NFMR), Theme Understanding Accuracy (NTUA), and Coherence Score (NCS). Our algorithm achieved NFMR = 89.7% (+37.58% over GA; +14.27% over IGA-Unconstrained),NTUA = 85.0%(+40.96% over GA;+17.08% over IGA-Unconstrained), and NCS = 7.95 (+29.27% over GA;+9.96% over IGA-Unconstrained). These gains confirm its strength in capturing narrative functions, interpreting themes, and sustaining logical coherence−better aligning with human cognition. By refining narrative-structure constraints and thematic-semantic mapping, it establishes a robust foundation for high-quality automated narrative generation (Table 2).

Table 2. Narrative Adaptation Evaluation Results (NFMR: Narrative Function Matching Rate; NTUA: Narrative Theme Understanding Accuracy; NCS: Narrative Coherence Score)

Algorithm	NFMR(%)	NTUA (%)	NCS(points)
GA	65.2	60.3	6.15
PSO	68.9	63.5	6.42
Greedy Algorithm	72.3	65.8	6.78
IGA-Unconstrained	78.5	72.6	7.23
Proposed Algorithm	89.7	85.0	7.95

- **Functional Constraint Satisfaction Evaluation**

Table 3 evaluates functional constraint satisfaction across three metrics: Building Spacing Compliance Rate, Traffic Accessibility Score, and Orientation Compatibility Score. Our algorithm achieves 96.8% spacing compliance (+14.3 percentage points over GA and +4.5 percentage points over IGA-Unconstrained), demonstrating strong adherence to spatial regulations. It scores 8.78 in traffic accessibility and 8.69 in orientation compatibility, outperforming the Greedy Algorithm by 0.89 and 0.97 points and PSO by 1.21 and 1.22 points, respectively. These results confirm its ability to jointly

optimize traffic efficiency and orientation suitability. Its consistent superiority over all baselines (GA, PSO, and IGA-Unconstrained) validates robustness for high-density urban layout optimization.

Table 3. Functional Constraint Satisfaction Evaluation Results

Algorithm	Bldg. Spacing Compl. Rate(%)	Traff. Access. Score(pts)	Orient. Compat. Score(pts)
GA	82.5	7.23	7.15
PSO	85.6	7.56	7.48
Greedy Algorithm	88.9	7.89	7.72
IGA-Unconstrained	92.3	8.25	8.16
Proposed Algorithm	96.8	8.78	8.69

- User Testing Results

Thirty participants meet the recommended sample size (25–30) for quantitative VR user experience studies to ensure statistical validity—including five literature enthusiasts, fifteen VR undergraduates, five landscape animation graduate students, and five VR researchers—evaluated the technology–humanity integration. Results show that 24 participants (80%) agreed that the VR scenes more effectively convey the magical realist aesthetic and science-fictional atmosphere portrayed in the original text of *Not In This Lifetime*; 23 participants (77%) noted strong sensory impact fused with the source text's aesthetic imagery; 26 participants (85%) rated the aesthetic quality and interactivity as satisfactory; and 22 participants (73%) found the VR experience more effective than traditional reading for grasping abstract sci-fi and magical realism imagery. Statistical analysis: The differences between the proposed method and all baselines are significant ($p < 0.05$, independent samples t-test). The inter-rater reliability of expert scores is Cronbach's $\alpha = 0.88$, indicating high evaluation consistency. The user study results show a 95% CI [7.52, 8.20] for the aesthetic score. A control group using conventional text-to-VR manual layout is employed for comparison; the proposed method outperforms the control group by 2.3 points in subjective aesthetic score and 31.6% in narrative function matching rate, verifying its superiority over traditional VR scene construction methods (Figure 3).



Figure 3. Optimized VR Scene Layouts Corresponding to *Not In This Lifetime* Chapters (Top left: panoramic view of the urban village corresponding to Chapters 5–7; top right: partial view of the urban village corresponding to Chapters 1–2; bottom left: city skyline view from Chapter 3; bottom right: vernacular market scene from Chapter 9.).

- Ablation Study

To verify the effectiveness of each core module and confirm performance gains arise from synergistic integration rather than narrative constraints alone, a controlled ablation study is conducted under identical experimental configurations. Five groups are designed: Baseline (IGA-Unconstrained), w/o Scale-adaptive Fusion, w/o ConstraintCheck, w/o Dynamic Feedback, and Full Model (proposed algorithm). Quantitative results are reported in Table 4, with statistical significance tested via independent-sample t-tests ($p < 0.05$ vs. Full Model).

Table 4. Ablation study results (Note: $p < 0.05$ vs. Full Model; NFMR: Narrative Function Matching Rate; NCS: Narrative Coherence Score; values are presented as mean ± standard deviation (SD))

Group	Setting	C_{rule3}	C_{golden}	NFMR (%)	NCS (pts)	Spacing Rate (%)
1	Baseline	0.823±0.041	0.795±0.038	78.5±2.4	7.23±0.31	92.3±1.6
2	w/o Scale-adaptive Fusion	0.835±0.037	0.808±0.035	80.6±2.2	7.41±0.29	93.4±1.5
3	w/o ConstraintCheck	0.851±0.034	0.824±0.032	83.2±2.0	7.60±0.27	94.5±1.3
4	w/o Dynamic Feedback	0.859±0.031	0.833±0.030	84.7±1.8	7.72±0.25	95.1±1.2
5	Full Model	0.887±0.028*	0.856±0.026*	89.7±1.5*	7.95±0.22*	96.8±1.0*

The Full Model significantly outperforms all ablated variants across all metrics ($p < 0.05$). Removing ConstraintCheck or Dynamic Feedback markedly reduces narrative coherence, while omitting scale-adaptive fusion degrades layout uniformity. Results demonstrate that performance improvements originate from the synergistic fusion of the proposed modules, validating the rationality and superiority of the full framework.

- Experimental Conclusions

The proposed full-chain interactive genetic algorithm with dynamic feedback and dual constraints is validated via comparative experiments, surpassing baseline methods in aesthetic quality, narrative adaptation, and functional compliance. Generated layouts conform better to aesthetic rules, achieve higher expert scores, integrate literary narrative logic to enhance adaptability, and satisfy functional constraints (Figure 4). User experiments verify effective tech-humanistic integration: the algorithm efficiently produces VR scenes consistent with the original text's theme and aesthetics, boosting user immersion and deepening literary understanding of the source material.



Figure 4. Cyberpunk-style virtual reconstruction of the ancient site of King Zhao's City based on the original text of *Not In This Lifetime* .

5. Conclusion and Future Work

Using the magical realist novel *Not In This Lifetime* as a testbed, this study proposes an algorithmic framework that unifies multimodal perception, aesthetic principles, and narrative logic [15]. It introduces an interactive genetic algorithm for large-scale 3D architectural layout, validates its effectiveness through empirical experiments, and examines the technology‑humanities relationship. Technically, the framework advances literary imagery visualization−key metrics (e.g., aesthetic scores) surpass those of traditional methods−and its full-chain variant outperforms baselines in aesthetic quality, narrative adaptation, and functional compliance. This study provides a new technical path for literary immersive visualization. Future work will explore balancing user personalization with literary aesthetics, build dynamic preference models, extend the algorithm to diverse literary VR applications, and integrate with diffusion-based text-to-3D models−advancing cross-media literary dissemination and intelligent VR production.

References

[1] Zhang, Y. (2024). Not in This Lifetime (science fiction work of magical realism_Chapters 1-10). Zenodo. https://doi.org/10.5281/zenodo.18058055

[2] Zhang Y, Yang G. Optimization of the Virtual Scene Layout Based on the Optimal 3D Viewpoint[J]. IEEE Access, 2022, 10(10):1-1. DOI:10.1109/ACCESS.2022.3214206.

[3] Bartolomeo P. The Attention Systems of the Human Brain[M]. Attention Disorders After Right Brain Damage. Springer London, 2014.

[4] Itti L, Koch C, Niebur E. A Model of Saliency-based Visual Attention for Rapid Scene Analysis[J]. In Proceedings of IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(11):1254-1259.

[5] Cho S B. Towards creative evolutionary systems with interactive genetic algorithm[J]. In Proceedings of Applied Intelligence, 2002, 16, pp:129-138.

[6] Muraoka N, Aoki Y. Optimization of building form by genetic algorithm which has learning process of the evaluation criterion[J]. In Journal of Architecture & Planning, 1998, 63(514), pp:141-146.

[7] Yang X, Hu F, Liu S, et al.3D indoor scene assessment via layout plausibility[J]. Displays, 2025, 87:102964. DOI:10.1016/j.displa.2025.102964.

[8] Cai Y .Optimization of landscape element layout in environmental art design based on genetic algorithm[J].JOURNAL OF COMPUTATIONAL METHODS IN SCIENCES AND ENGINEERING, 2025, 25(5):4246–4259.DOI:10.1177/14727978251337997.

[9] Fan R, Wang L, Liu X, et al. Real-scene-constrained virtual scene layout synthesis for mixed reality[J]. The Visual Computer, 2024, 40(9):21.DOI:10.1007/s00371-023-03167-4.

[10] Wu B S, Chen H E, Huang S Y, et al. TPA3D: Triplane Attention forFast Text-to-3D Generation[C]//European Conference on Computer Vision. Springer, Cham, 2025.DOI:10.1007/978-3-031-72649-1_25.

[11] Lee H , Savva M , Chang A X .Text‐to‐3D Shape Generation[J].Computer Graphics Forum, 2024, 43(2).DOI:10.1111/cgf.15061.

[12] Ma Y, Fan Y, Ji J, et al. Creating High-quality 3D Content by Bridging the Gap Between Text-to-2D and Text-to-3D Generation[J].ACM Transactions on Multimedia Computing, Communications, and Applications, 2024.DOI:10.1145/3687475.

[13] Zhang, Y. (2024). Not in This Lifetime (science fiction work of magical realism_Chapters 11-20). Zenodo. https://doi.org/10.5281/zenodo.19549506

[14] Knight E, Balzter H, Breeze TD, et al. Adapting genetic algorithms for multifunctional landscape decisions: A theoretical case study on wild bees and farmers in the UK[J]. Methods in Ecology & Evolution, 2024, 15(11). DOI:10.1111/2041-210X.14424.

[15] ZHANG Y. Not in This Lifetime (此生无望) [M]. Singapore: Universal Publishing House Pte. Ltd., 2024. ISBN 978-981-5271-72-0: pp. 430‑583.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)

© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261101

Lightweight Face Anti-Spoofing for Mobile Multi-Factor Attendance Systems

Ekkachan Rattanalerdnusorn[a,1] and Sasakorn Pichetjamroen[a]
[a]*National Electronics and Computer Technology Center (NECTEC), National Science and Technology Development Agency (NSTDA), Pathum Thani, Thailand*
ORCiD ID: Ekkachan Rattanalerdnusorn https://orcid.org/0000-0002-0783-2059

Abstract. Mobile applications using facial verification are increasingly adopted across government services, financial platforms, and organizational access control in Thailand, where secure and reliable authentication is critical. However, such systems remain vulnerable to presentation attacks, including photo spoofing and replay attacks, which undermine trust in biometric authentication. This paper presents a lightweight AI-enabled multi-factor attendance system (*AtTime*) with integrated face anti-spoofing for secure mobile authentication. Unlike previous studies that primarily focused on photo-based spoof detection under controlled conditions, the proposed approach utilizes facial images captured directly from smartphone front-facing cameras and integrates anti-spoofing into a real-world mobile attendance platform. The proposed framework combines facial verification and anti-spoofing within a practical system architecture without requiring additional hardware, enabling cost-effective deployment in mobile environments. The anti-spoofing module, based on EfficientNet-B0, was trained on 12,032 images and validated on 1,860 images. Field experiments involving 175 participants achieved a spoof detection rate of 97.20% and an overall F1-score of 90.06% while maintaining usability. Furthermore, the results indicate that an operational distance of 10–15 cm improves spoof detection performance to 96.19%. Although the evaluation focused primarily on smartphone-based photo spoofing scenarios, the results demonstrate that integrating lightweight AI-based anti-spoofing into mobile authentication systems can effectively enhance security, user trust, and practical deployment in real-world digital transformation environments.

Keywords. Multi-Factor Authentication, Face Anti-Spoofing, Facial Verification, Mobile Authentication, Biometric Security, Deep Learning.

1. Introduction

Facial verification is widely used in modern mobile applications due to its convenience, contactless operation, and ease of use. As facial authentication becomes increasingly adopted in digital services, ensuring its reliability and security has become a critical requirement [1], [2]. However, facial authentication systems remain vulnerable to presentation attacks, particularly photo spoofing and replay attacks, where attackers use printed images or displayed facial media to bypass authentication [3], [4]. Furthermore, emerging threats such as Deepfake technologies introduce increasingly sophisticated attack scenarios, making it more challenging to distinguish between live and spoof facial inputs [5]. Existing approaches to mitigate these threats often rely on challenge-response

[1] Corresponding Author: Ekkachan Rattanalerdnusorn, Email: Ekkachan.rattanalerdnusorn@nectec.or.th

techniques, such as blinking detection or head movement verification [6]. While these methods can improve security, they may reduce usability and often require additional user interaction. Moreover, some anti-spoofing approaches depend on specialized hardware, such as depth sensors, increasing system cost and limiting practical deployment in mobile environments [7].

Previous studies on the *AtTime* platform investigated photo-based face anti-spoofing using several CNN architectures, including MobileNet, NASNet Large, and EfficientNet [8]. Although these studies demonstrated promising performance, they primarily focused on photo spoofing under controlled conditions. In practical mobile authentication scenarios, presentation attacks using facial images displayed on smartphone screens are more realistic because attackers can directly use mobile devices without requiring printed photographs. However, the effectiveness of anti-spoofing approaches against smartphone-based presentation attacks in real-world operational environments remains insufficiently investigated, **representing an important research gap for practical mobile authentication systems.**

To address these challenges, this paper presents a lightweight AI-enabled multi-factor authentication framework implemented through a real-world mobile attendance system, namely *AtTime*. The proposed system integrates facial verification with a face anti-spoofing mechanism based on EfficientNet-B0 [9]. EfficientNet-B0 was selected because it provides a balance between detection performance and model complexity, making it suitable for real-world mobile deployment. Unlike previous *AtTime* studies that focused primarily on photo-based spoof detection, the proposed framework specifically investigates smartphone-based presentation attacks and utilizes facial images captured directly from smartphone front-facing cameras under actual operational conditions. This design eliminates the need for additional hardware while maintaining low computational cost, enabling reliable detection of live and spoof facial images using only two-dimensional inputs.

The main contributions of this paper are as follows: (1) a lightweight face anti-spoofing method using two-dimensional facial images without additional hardware; (2) a practical anti-spoofing framework for smartphone-based presentation attacks; (3) integration into a real-world multi-factor attendance system (AtTime); and (4) experimental validation involving 175 participants under operational conditions.

2. Related Work

The *AtTime* platform is a mobile attendance system previously developed by the research team, supporting multi-factor authentication and facial verification across mobile and kiosk-based environments [1], [2]. While the platform improves attendance security and user convenience, it remains vulnerable to presentation attacks, particularly photo spoofing, where printed facial images may be used to bypass authentication. This limitation motivates the integration of face anti-spoofing techniques in the present study.

Presentation Attack Detection (PAD) plays a critical role in protecting biometric authentication systems from spoofing attacks [7], [10]. According to ISO/IEC 30107-3:2023, common presentation attacks include photo attacks, replay attacks, mask attacks, and Deepfake-based impersonation [3], [11], [12]. Existing PAD approaches often rely on challenge-response mechanisms or specialized hardware, which may reduce usability, increase deployment cost, and limit practicality in mobile environments [6], [7], [10].

Recent advances in deep learning have significantly improved face anti-spoofing performance. CNN-based architectures, such as MobileNet, NASNet, and EfficientNet, have been widely adopted for presentation attack detection [10]. Previous work on the *AtTime* platform compared these architectures and demonstrated the effectiveness of CNN-based approaches for spoof detection in attendance systems [8]. Among them, EfficientNet achieved the best performance while maintaining computational efficiency through its compound scaling strategy [8], [9].

Recent studies have explored advanced deep learning architectures for face anti-spoofing and presentation attack detection. Transformer-based approaches, such as Vision Transformers (ViT) [10], have demonstrated promising performance by capturing global feature relationships and improving generalization across diverse attack scenarios. In addition, temporal-based approaches, including CNN-LSTM architectures [13], [14], combine spatial and temporal information to detect replay attacks and manipulated facial content. These methods have shown strong performance in video-based anti-spoofing and Deepfake detection tasks. The rapid advancement of Deepfake generation techniques has further increased the difficulty of distinguishing genuine facial inputs from synthetic content, creating new challenges posing new challenges to biometric authentication systems [15], [16].

Despite their effectiveness, transformer-based and temporal-based approaches generally require larger datasets, higher computational resources, and additional processing overhead, which may limit their practicality for lightweight mobile applications. Therefore, this study adopts EfficientNet-B0 as the backbone architecture because it provides a balance between detection performance and model complexity while maintaining deployment simplicity for real-world mobile attendance systems.

Table 1. Comparison of Recent Face Anti-Spoofing Approaches.

Method	Architecture	Strengths	Limitations
Vision Transformer (ViT)	Transformer	Captures global feature relationships and improves generalization	High computational complexity
CNN-LSTM	Hybrid CNN + Temporal Modeling	Effective for replay attack and temporal analysis	Requires sequential image frames
Deepfake Detection Models	CNN / Transformer / Hybrid	Effective against AI-generated facial content	Requires specialized datasets and larger computational resources
EfficientNet-B7	CNN	High detection accuracy	Large model size and computational cost
Proposed EfficientNet-B0	CNN	Lightweight and suitable for mobile deployment	Limited evaluation on advanced attack types

3. Design Methodology

3.1. System Architecture

This section describes the system architecture, anti-spoofing process, dataset acquisition, image preprocessing, and model development of the proposed mobile attendance system.

Figure 1(a) illustrates the architecture of the proposed *AtTime* mobile attendance platform deployed in a cloud-based environment. Users access the system through a smartphone application connected to backend services via an API Gateway. Authentication, request validation, and secure communication are performed before

facial images are transmitted for face detection and anti-spoofing analysis. To enhance security, the platform employs session-based API keys together with RSA and AES encryption to mitigate replay attacks and unauthorized access.

After authentication, facial images are forwarded to the Image and Model Service for face detection. If a valid face is detected, the image is processed by the Anti-Spoofing Service, where the trained model classifies the input as either Live or Spoof. The classification result is subsequently used for attendance validation.

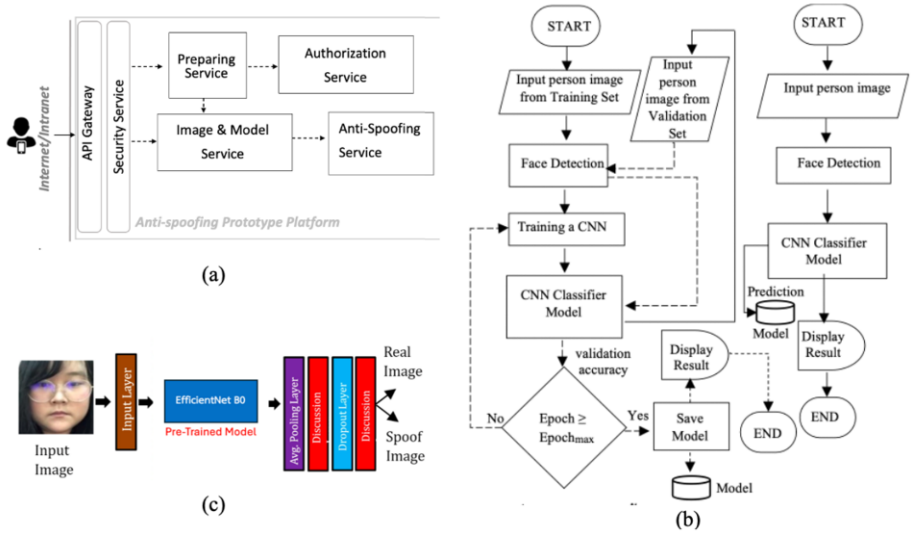


Figure 1. (a) Mobile attendance system architecture; (b) Anti-spoofing process; (c) Layer architecture of the proposed face anti-spoofing model.

3.2. Anti-Spoofing Framework

The anti-spoofing framework consists of training and testing stages, as illustrated in Figure 1(b). During training, facial images are divided into training and validation sets, processed through face detection, and used to train the CNN-based model. The best-performing model is selected based on validation performance.

During testing, the trained model is integrated into the *AtTime* platform for real-time classification of facial images captured from mobile devices. The framework prioritizes spoof detection to reduce the risk of unauthorized access in mobile attendance systems.

3.3. Dataset Acquisition

Unlike our previous photo-based anti-spoofing study, which relied on facial photographs collected under controlled conditions, the proposed dataset acquisition framework utilizes images captured directly from smartphone front-facing cameras during real-world operation. As illustrated in Figure 2, live samples were collected from legitimate users through the AtTime mobile application, while spoof samples were generated using facial images displayed on smartphone screens under screen flash illumination. In this study, the term "screen flash" refers to white screen illumination generated by the smartphone display during image acquisition rather than the device's camera flash. This illumination was designed to support image capture by introducing additional visual cues

for anti-spoofing analysis while reducing potential visual discomfort during use. This acquisition strategy was designed to reflect practical mobile authentication scenarios and realistic presentation attack conditions.

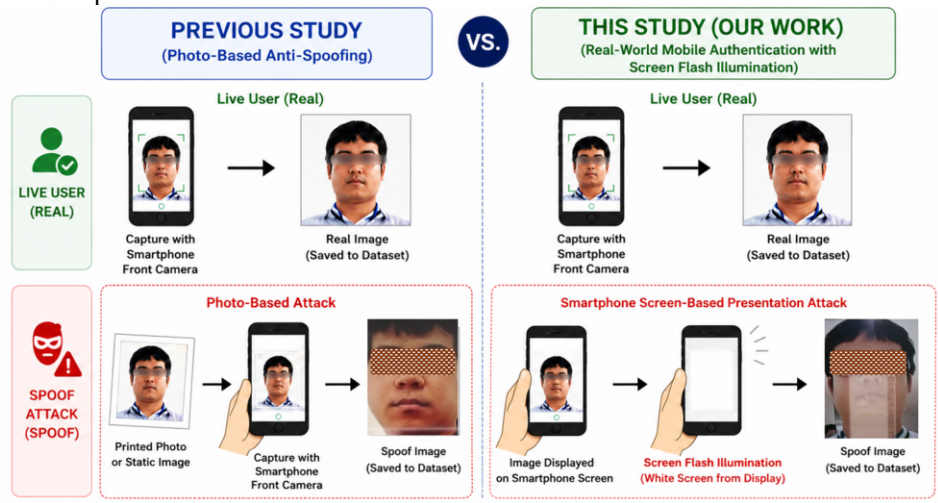


Figure 2. Comparison of Photo-Based and Smartphone Screen-Based Dataset Acquisition Frameworks.

The dataset consists of 6,010 live images and 6,022 spoof images for training, while the validation set includes 880 live images and 980 spoof images. A balanced distribution of live and spoof samples was maintained to reduce classification bias during training.

Table 2. Distribution of Live and Spoof Images in the Dataset.

Dataset	Live	Spoof	Total
Training	6,010	6,022	12,032
Validation	880	980	1,860
Total	6,890	7,002	13,892

3.4. Image Preprocessing

All facial images were resized to 224 × 224 pixels and processed as RGB images before training. Image normalization was applied using TensorFlow ImageDataGenerator. No explicit data augmentation was performed because the dataset was collected under operational conditions and was intended to reflect real-world mobile attendance scenarios.

3.5. EfficientNet-B0 Anti-Spoofing Model

The proposed anti-spoofing model is based on EfficientNet-B0, which was selected for its balance between detection performance and computational efficiency. The original classification layer was replaced with a custom head consisting of a Global Average Pooling layer, Dense layer, Dropout layer, and a final Dense layer for binary classification (Live and Spoof), as illustrated in Figure 1(c).

The model was trained using the prepared dataset, and the best-performing model was selected based on validation performance. The final model contains approximately 5.36 million parameters, making it suitable for deployment in resource-constrained mobile authentication environments. The model configuration and training hyperparameters are summarized in Table 3.

Table 3. Hyperparameters and Model Configuration.

Parameter	Value
Backbone Network	EfficientNet-B0
Input Resolution	224 × 224 × 3
Optimizer	Adam
Learning Rate	0.00001
Batch Size	256
Epochs	30
Dropout Rate	0.6
Output Activation	Sigmoid
Number of Classes	2
Total Parameters	5,363,365
Trainable Parameters	5,321,342

4. Experimental Results

The evaluation was conducted at the CIC Skill Development and Learning Center, Phuket, Thailand, during March 31–April 6, 2024. A total of 175 participants were involved to simulate real-world mobile attendance scenarios.

4.1. Experimental Setup

The evaluation was conducted at the CIC Skill Development and Learning Center, Phuket, Thailand, during March 31–April 6, 2024. A total of 175 participants were involved to simulate real-world mobile attendance scenarios.

Two mobile devices were used: a Samsung Galaxy Tab S3 (9.7-inch display) and a Motorola Moto G5 Plus (5.2-inch display), both equipped with 5 MP front cameras. Live face testing was performed at capture distances of 10–15 cm, 30 cm, and 40 cm to evaluate the impact of facial capture distance on anti-spoofing performance. For spoof testing, facial images displayed on smartphone screens were recaptured using the front-facing camera under screen flash illumination to simulate smartphone screen-based presentation attacks.

Model performance was evaluated under different devices and capture distances using detection accuracy. The corresponding results are summarized in Table 4.

Table 4. Detection Accuracy for Live and Spoof Samples across Different Devices and Capture Distances.

Testing Mobile	Confusion Matrix					Accuracy (%)
	Actually Class	Face distance/ Device	Number of Testing	Predicted Class		
				Live	Spoof	
Samsung	Live	10-15 cm	175	130	45	74.29%
		30 cm	175	125	50	71.43%
		40 cm	175	124	51	70.86%
	Spoof	10 -15 cm with Motorola	365	18	347	95.07%
		10 -15 cm with Samsung	351	9	342	97.44%
Motorola	Live	10-15 cm	175	150	25	85.71%
		30 cm	175	127	48	72.57%
		40 cm	175	128	47	73.14%
	Spoof	10 -15 cm with Motorola	356	4	352	98.88%
		10 -15 cm with Samsung	355	9	346	97.46%

4.2. Performance Evaluation

The proposed EfficientNet-B0-based anti-spoofing model was evaluated using a total of 2,477 test cases. As shown in Table 5 (left), the model correctly classified 1,387 spoof images and 784 live images, while misclassifying 40 spoof images as live and 266 live images as spoof. To further investigate the effect of capture distance, an additional evaluation was conducted under close-range conditions (10–15 cm). The corresponding confusion matrix is presented in Table 5 (right).

Table 5. Summary of detection performance across all conditions and under close-range conditions(10–15 cm).

Number of Test Cases: 2,477		**All test conditions**		**Number of Test Cases: 1,777**		**Close-range (10–15 cm)**	
		Spoof (+)	Live (-)			Spoof (+)	Live (-)
Predicted	Spoof (+)	1387	40	Predicted	Spoof (+)	1387	40
	Live (-)	266	784		Live (-)	70	280

The system achieved a True Positive Rate (spoof detection) of 97.20% and a True Negative Rate (live detection) of 74.67%, resulting in an overall accuracy of 87.65%, as presented in Table 6. Under close-range conditions (10–15 cm), the model achieved an F1-score of 96.19%, indicating that shorter capture distances improve spoof detection performance.

The results indicate that the proposed model is more conservative in classifying live users, resulting in a lower live-face detection rate compared with spoof detection performance. This behavior is desirable in attendance authentication systems because it minimizes the risk of false acceptance, where spoof attacks are incorrectly classified as legitimate users. Consequently, the framework prioritizes security and spoof prevention over user convenience.

Table 6. Detection accuracy for live and spoof facial images.

Type of Evaluation	**Accuracy (100%)**
Live Face detection	74.67%
Spoof Face Detection	97.20%
Overall detection (Live and Spoof)	87.65%

Table 7 presents the performance breakdown across different indoor testing environments. Under the controlled indoor setup using the *AtTime* system, the Bangkok cohort (N = 84) achieved an accuracy of 96.43% with a False Rejection Rate (FRR) of 3.57%. In contrast, the Phuket cohort (N = 175) achieved an accuracy of 74.29% with an FRR of 25.71%. Although both evaluations were conducted indoors, differences in lighting conditions, environmental settings, and participant characteristics may have contributed to the observed performance variation. These findings suggest that environmental and demographic factors can influence anti-spoofing performance and should be investigated further in future studies.

Table 7. Performance Breakdown Under Different Indoor Testing Conditions.

Evaluated Condition	Sample Size (N)	Accuracy (%)	False Rejection Rate (FRR %)	**Accuracy (100%)**
Bangkok Cohort (Controlled Indoor Setup)	84	81	3.37%	96.43%
Southern Cohort (Phuket Indoor Setup)	175	130	25.72%	74.28%

As shown in Figure 3, the proposed EfficientNet-B0 model achieved an Area Under the ROC Curve (AUC) of 0.9485, demonstrating strong discriminative capability between live and spoof facial images. This result is consistent with the high spoof detection rate and F1-score reported in Table 6.

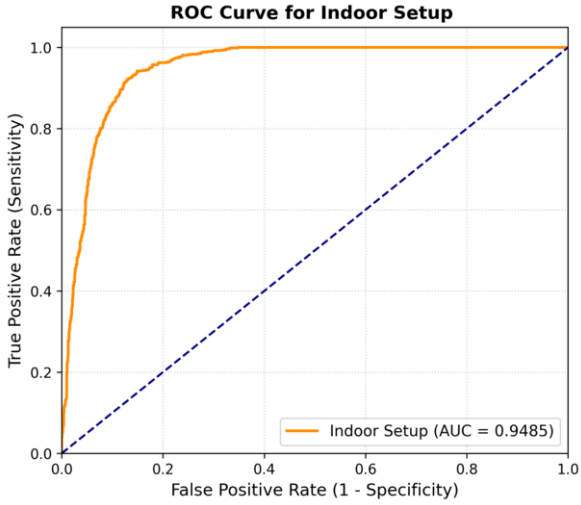


Figure 3. Receiver Operating Characteristic (ROC) curve of the proposed EfficientNet-B0 model for live and spoof classification (AUC = 0.9485).

4.3. Distance-Based Performance Analysis

To further evaluate the impact of capture distance, additional experiments were conducted under close-range conditions (10–15 cm), as shown in Table 4 (right). From 1,777 test cases, the proposed model achieved a Precision of 97.20%, a Recall of 95.20%, and an F1-score of 96.19%, with an overall accuracy of 93.81%.

Compared with the overall evaluation results, performance improved under close-range conditions, indicating that capture distance influences anti-spoofing effectiveness. These results suggest that maintaining a facial capture distance of approximately 10–15 cm can improve detection reliability in practical mobile authentication scenarios.

5. Discussion

Compared with previous studies on the AtTime platform, the proposed EfficientNet-B0 model provides a favorable balance between model complexity and anti-spoofing performance. With approximately 5.36 million parameters, the model is substantially lighter than larger CNN architectures while remaining suitable for deployment in resource-constrained mobile authentication environments.

The experimental results demonstrate strong spoof detection performance, achieving a True Positive Rate of 97.20% and an F1-score of 90.06%. Although the True Negative Rate for live users is lower at 74.67%, this behavior reflects a security-oriented design that prioritizes spoof prevention over user convenience. In practical deployment, false rejection can be mitigated through re-verification, whereas false acceptance may result in unauthorized access. The relatively high False Rejection Rate (FRR) may be attributed to variations in capture distance, lighting conditions, facial pose, and image quality during real-world operation. These factors can reduce facial detail visibility and affect

the extraction of discriminative features, causing some live samples to be incorrectly classified as spoof.

The results further indicate that capture distance influences anti-spoofing performance. Under close-range conditions (10–15 cm), the model achieved an improved F1-score of 96.19%, suggesting that shorter capture distances provide more discriminative facial information for spoof detection. This finding further suggests that insufficient facial detail at longer capture distances may contribute to false rejection cases. This finding offers practical guidance for real-world mobile authentication deployment.

A key advantage of the proposed framework is its practical deployment capability. Unlike approaches requiring specialized hardware such as depth sensors, the system operates using only two-dimensional facial images captured from smartphone front-facing cameras. This reduces deployment cost and enables integration into existing mobile attendance platforms.

Despite these promising results, several limitations remain. The dataset contains limited diversity in lighting conditions, device variations, and participant demographics. Furthermore, no dedicated fairness evaluation was conducted across demographic groups such as different skin tones, age groups, or genders. Performance variation observed across the Bangkok and Phuket cohorts suggests that environmental conditions and participant diversity may influence anti-spoofing performance. Future work should incorporate more diverse datasets and controlled evaluations to improve robustness, generalization performance, and fairness assessment.

In addition, this study focuses on photo spoofing and smartphone screen-based presentation attacks. More advanced threats, including video-based replay attacks, AI-generated Deepfake attacks, and recent transformer-based or temporal-based anti-spoofing approaches, require further investigation to enhance the robustness of future Presentation Attack Detection (PAD) systems.

6. Conclusion

This paper presents an EfficientNet-B0-based face anti-spoofing framework for secure mobile attendance systems using two-dimensional facial images captured from smartphone front-facing cameras. The proposed framework was integrated into the AtTime platform and evaluated through real-world testing involving 175 participants.

Experimental results demonstrate the effectiveness of lightweight AI-based anti-spoofing for mobile authentication under practical deployment scenarios. The findings further indicate that capture distance influences spoof detection performance, providing useful guidance for real-world mobile attendance applications.

The proposed framework operates without specialized hardware and incorporates session-based API keys together with RSA/AES encrypted communication to enhance authentication security. These characteristics enable cost-effective deployment and integration into existing mobile authentication platforms.

Despite these promising results, the current study is limited by the diversity of environmental conditions, device variations, and participant demographics, which may affect model generalization across different deployment scenarios.

Future work will focus on expanding dataset diversity and incorporating more advanced Presentation Attack Detection (PAD) mechanisms, including video-based replay attacks, AI-generated Deepfake detection, 3D mask attacks, and transformer-

based or temporal-based anti-spoofing approaches to further improve robustness, fairness, and security in mobile authentication systems.

Acknowledgment

This work was supported by the Fundamental Fund under the Science, Research and Innovation Promotion Fund. The authors thank the Phuket City Municipality and participants at the CIC Skill Development and Learning Center for their support during the field experiments.

References

[1] Jaikla T, Pichetjamroen S, Vorakulpipat C, Pichetjamroen A. A Secure Four-Factor Attendance System for Smartphone Device. In: *2020 22nd International Conference on Advanced Communication Technology (ICACT)*; 2020 Feb 16-19; Phoenix Park, Korea. p. 65-68, doi: 10.23919/ICACT48636.2020.9061545

[2] Pichetjamroen S, Rattanalerdnusorn E, Vorakulpipat C, Pichetjamroen A. Multi-Factor Based Face Validation Attendance System with Contactless Design in Training Event. In: *2021 18th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*; 2021 May 19-22; Chiang Mai, Thailand. p. 637-640, doi: 10.1109/ECTI-CON51831.2021.9454817

[3] Yu Z, Zhao C, Lei Z. Face Presentation Attack Detection. *arXiv:2212.03680 [Preprint]*. 2022, doi: 10.48550/arXiv.2212.03680

[4] Smith DF, Wiliem A, Lovell BC. Binary Watermarks: A Practical Method to Address Face Recognition Replay Attacks on Consumer Mobile Devices. In: *2015 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA)*; 2015 Mar 23-25; Hong Kong, China. p. 1-8, doi: 10.1109/ISBA.2015.7126359.

[5] Yu Z, Li X, Zhao G. Deep Learning for Face Anti-Spoofing: A Survey. *IEEE Trans Pattern Anal Mach Intell.* 2023, doi: 10.1109/TPAMI.2023.3276182

[6] Anjith George A, Marcel S. Deep Pixel-Wise Binary Supervision for Face Presentation Attack Detection. In: *2019 International Conference on Biometrics (ICB)*; 2019 Jun 4-7; Crete, Greece. p. 1-8, doi: 10.1109/ICB45274.2019.8987369

[7] Hernandez-Ortega J, Fierrez J, Morales A, Galbally J. Introduction to Face Presentation Attack Detection. In: *Handbook of Biometric Anti-Spoofing*. 2nd ed. Cham (CH): Springer; 2019. p. 103-142, doi: 10.1007/978-3-030-01931-0_5

[8] Pichetjamroen S, Rattanalerdnusorn E, Lee-Laruji T, Vorakulpipat C. A Comparative Study of 2D Image Anti-Spoofing Classification on Time Attendance Platform. In: *2022 8th International Conference on Computing and Artificial Intelligence (ICCAI '22)*; 2022 Mar 18-21; Tianjin, China. p. 546-551, doi: 10.1145/3532215.3532241

[9] Tan M, Le Q. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In: *Proceedings of the 36th International Conference on Machine Learning (ICML)*; 2019 Jun 9-15; Long Beach, CA. p. 6105-6114.

[10] Yu Z, Zhao C, Lei Z. Deep Learning for Face Anti-Spoofing: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2023;45(5):5609–5631.

[11] International Organization for Standardization. *ISO/IEC 30107-3:2023, Information Technology — Biometric Presentation Attack Detection — Part 3: Testing and Reporting*. Geneva: ISO; 2023.

[12] Gernot T, Plateaux P, Rosenberger C. Robust Biometric Scheme Against Replay Attacks Using One-Time Biometric Templates. *Comput Secur.* 2024 Mar;140:103762, doi: 10.1016/j.cose.2024.103762

[13] A Hybrid CNN-LSTM Approach for Precision Deepfake Image Detection Based on Transfer Learning. Electronics. 2024;13(9):1662.

[14] An Investigation into the Utilisation of CNN with LSTM for Video Deepfake Detection. 2025.

[15] Passos LA, Jodas D, Costa KAP, et al. A Review of Deep Learning-Based Approaches for Deepfake Content Detection. Expert Systems. 2024;41(8):e13570.

[16] Awan MJ, et al. A Contemporary Survey on Deepfake Detection: Datasets, Algorithms, and Challenges. Electronics. 2024;13(3):585.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)

© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261102

Hybrid Financial Analysis Agent with RAG and Text Mining

Bingzhang WANG [a], Keyi ZHANG [a], Tianhao DU [a], Xiaojie CHEN [a], Jiyu CAO [a], Shaokang HE and Adela S.M. LAU [b, 1]

[a] *Department of Statistics and Actuarial Science, School of Computing and Data Science, The University of Hong Kong, China*

[b] *Data Science Lab, Department of Statistics and Actuarial Science, School of Computing and Data Science, The University of Hong Kong, China*

ORCiD ID:
Bingzhang Wang https://orcid.org/0009-0009-5452-4209
Keyi Zhang https://orcid.org/0009-0004-8896-6201
Tianhao Du https://orcid.org/0009-0009-7238-0683
Xiaojie Chen https://orcid.org/0009-0007-8886-4004
Jiyu Cao https://orcid.org/0009-0003-0111-8107
Shaokang He https://orcid.org/0009-0009-6667-7835
Adela S.M. Lau https://orcid.org/0000-0001-5918-8309

Abstract. This paper presents a hybrid financial analysis agent that integrates retrieval-augmented generation, financial text mining, live market-data access, and a news sentiment classifier within a LangGraph-based orchestration framework. The system is designed to reduce information overload for everyday investors by routing natural-language financial questions to domain knowledge retrieval, market data, company fundamentals, web/news evidence, and sentiment analysis. Experimental results on the FinancialPhraseBank AllAgree subset show that a TF-IDF and LinearSVC sentiment classifier achieves strong prototype-level performance, while the integrated demo demonstrates structured, evidence-aware financial question answering. The objective is to support user-level financial information access and interpretation rather than to claim direct causal improvement of market efficiency.

Keywords. Financial analysis agent, retrieval-augmented generation, text mining, sentiment analysis, market data

1. Introduction

1.1. Project background

Modern financial information is highly fragmented across multiple sources [1]. To make informed decisions, investors must simultaneously monitor and interpret fast-moving news, price movements, company fundamentals, and specialized terminology [2][3].

[1] Corresponding Author: Adela S.M. Lau, Data Science Lab, Department of Statistics and Actuarial Science, School of Computing and Data Science, Run Run Shaw Building, The University of Hong Kong, Pokfulam Road, Hong Kong; E-mail: adelalau@hku.hk.

1.2. Problem statement

This fragmentation creates severe information overload and significant domain knowledge barriers for ordinary users [4]. Integrating these isolated signals into a clear investment judgment requires high cognitive effort for everyday analysis, making the financial markets difficult to navigate without professional expertise [5][6].

1.3. Project aims and objectives

The primary aim of this project is to build an intelligent assistant for financial information understanding and analysis that lowers the barrier to entry for everyday investors. The core objectives are to:

- Perform Financial Q&A: Answer finance-related user questions accurately.
- Enable News Understanding: Interpret financial news and classify market sentiment.
- Execute Market Data Queries: Retrieve real-time price and trend information.
- Conduct Fundamental Analysis: Provide key company financial metrics, such as P/E ratios, EPS, and Market Cap.

1.4. Scope of study

The scope of this study encompasses the technical development and integration of the assistant's backend architecture. This includes connecting external data APIs (Twelve Data and Alpha Vantage), developing a hybrid Retrieval-Augmented Generation [7] pipeline combining dense and sparse search, and training classical machine learning models on the FinancialPhraseBank dataset to classify sentiment. The focus is on building an informational tool, not an automated trading system.

1.5. Proposed solution overview

Our proposed solution is an autonomous Large Language Model (LLM) agent architecture implemented through LangGraph-based workflow orchestration. An LLM planner dynamically routes user queries to the appropriate tools: a hybrid RAG system using Reciprocal Rank Fusion, external API modules for real-time market data, or a custom text mining component. Specifically, we deploy a TF-IDF and LinearSVC model to evaluate news sentiment. The agent synthesizes the outputs from these specialized modules to generate structured, evidence-supported financial insights for decision support.

2. Literature Review

2.1. Financial Information Analysis and Investor Decision Support

2.1.1. The Challenge of Financial Decision-Making

Financial investment decisions require synthesizing diverse information sources: real-time market data, company fundamentals, news events, and domain-specific knowledge [8]. Retail and institutional investors alike face significant challenges in this domain. Financial markets generate enormous amounts of data daily—stock prices, analyst notes, and news articles—making manual analysis time-consuming and error-prone. Financial metrics and terminology require domain knowledge, which increases the difficulty for

non-expert users [9]. Although some researchers used numerical data and textual data to analyze the stock price for investment decision [10][11], investment decisions are rarely based on a single data source, as investors need to correlate market trends, company performance, sentiment from news coverage, and long-term fundamentals, yet integrating these signals manually remains challenging. While general-purpose language models can provide investment advice, they often cannot cite evidence or justify recommendations, making it difficult for users to trust or validate the guidance [7].

2.1.2. Current Solutions and Their Limitations

Traditional financial information analysis relies on manual browsing of financial websites (e.g., stock information websites or market portals), where users search and read news one by one—an inefficient and subjective process. General chatbots (e.g., large language models like ChatGPT) can answer financial questions, but they often suffer from insufficient domain accuracy, outdated information, or factual hallucinations [7]. Professional financial data terminals (e.g., institutional-level analytics platforms) provide organized data and analytical tools; however, they are expensive, require specialized training, and remain inaccessible to most retail investors. Furthermore, some dedicated financial question-answering systems attempt to integrate market APIs and knowledge bases, but they tend to be fragmented due to the disjointed nature of market data, news sentiment, and knowledge retrieval.

2.1.3. Motivation for an Integrated System

To address the above limitations, a more sophisticated solution is required: an integrated financial decision-support assistant that retrieves and synthesizes multiple data modalities, including real-time and historical market data (prices, volume, fundamentals), sentiment from news articles and financial reports, and domain-specific knowledge covering financial terminology and common metrics. Such a system should provide grounded, explainable answers backed by retrieved evidence from documents, data points, and sources [7][12], allowing users to trace the reasoning and validate recommendations.

2.1.4. Role of Technology in This Project

This project implements the integrated system using three core technologies: Retrieval-Augmented Generation (RAG) [7] for financial QA based on a curated knowledge base, market and fundamental data APIs for real-time company metrics, and news sentiment analysis [13] to classify market sentiment. A key innovation of this project is the combination of traditional data science methods (e.g., TF-IDF based retrieval, supervised sentiment classification) with modern LLM-based agent techniques — the former providing interpretable, task-specific signals, and the latter orchestrating multi-source information to generate flexible responses. These components work together within a LangGraph-based agent framework that routes questions to appropriate tools and synthesizes outputs. Each technology will be detailed in the following sections.

2.2. Financial Question Answering Systems

Common architectures for financial QA systems include rule-based or template-based systems (handling fixed question types but poorly scalable), retrieval-based systems (retrieving relevant documents before answer generation, suitable for knowledge base queries [14]), and generation-based systems built on pretrained language models

(offering strong language understanding but prone to hallucinations). This project implements a RAG-based financial QA system [7] that combines retrieval and generation, preserving the traceability of knowledge base answers while improving the ability to handle open-ended natural language questions.

2.3. Retrieval-Augmented Generation for Domain-Specific QA

Retrieval-Augmented Generation (RAG) is an approach that combines information retrieval with language generation in a question-answering pipeline [7]. Instead of relying only on the internal knowledge of a language model, RAG first retrieves relevant external knowledge from a knowledge base and then uses the retrieved evidence to support the final generated answer. The general workflow of RAG includes two main stages: retrieving relevant knowledge snippets for the user query, and feeding both the original query and the retrieved context into a generation model to produce the final response [7].

This design is especially useful for domain-specific question answering. In specialized domains such as finance, many user questions involve professional terminology, financial indicators, abbreviations, and concept-level explanations. A general-purpose language model may not always provide sufficiently accurate or traceable answers for these questions. By introducing a finance-specific knowledge base, RAG allows the system to ground its responses in curated domain knowledge, such as financial definitions, common investor questions, and explanations of market-related concepts [12]. Therefore, RAG can improve both domain adaptation and factual reliability by synergistically merging the language model's intrinsic capabilities with dynamic, external databases [15].

One important benefit of RAG is explainability. Since the final answer is generated based on retrieved knowledge snippets, the system can provide source information or evidence that supports the response [7]. This is important in financial question answering because users often need to understand not only the answer itself, but also the basis for the explanation. Compared with a pure black-box chatbot, a RAG-based system can make the answer generation process more transparent and evidence-based.

Another advantage of RAG is its potential to reduce hallucination. In financial applications, unsupported or inaccurate answers may mislead users. By conditioning the generation model on retrieved domain-specific evidence, RAG reduces the likelihood that the model generates responses that are not grounded in the available knowledge base [7]. This makes it suitable for financial concept explanation, terminology clarification, and FAQ-style question answering.

However, the effectiveness of RAG depends heavily on retrieval quality. If the retrieved context is irrelevant or incomplete, the final generated answer may also be inaccurate. For this reason, practical RAG systems often adopt hybrid retrieval strategies that combine dense vector retrieval and sparse keyword retrieval [12][14]. Dense retrieval is useful for capturing semantic similarity between the user query and knowledge base content, while sparse retrieval methods such as BM25 are effective for matching exact terms, abbreviations, and specific financial expressions [16]. In financial QA, this combination is particularly valuable because users may ask semantically diverse questions, but many key concepts still depend on exact terminology.

Reranking is also commonly used as a second-stage retrieval technique. After the initial retrieval step obtains a set of candidate passages, a reranking method reorders these candidates according to their relevance to the query [14][17]. This follows the classic first-stage retrieval and second-stage reranking framework. The initial retrieval stage focuses on collecting potentially relevant results, while the reranking stage improves the precision of the final context. In domain-specific QA, reranking helps ensure that the most relevant financial knowledge is passed to the language model for answer generation.

Therefore, RAG is suitable for the financial QA component of this project, where concept-level answers require both domain adaptation and evidence grounding.

2.4. Financial News Sentiment Analysis

Financial news sentiment analysis aims to classify financial texts into positive, neutral, or negative sentiment categories, transforming unstructured news into structured signals for market interpretation [18]. Financial markets are highly sensitive to information, and breaking news, corporate announcements, earnings reports, and policy changes may influence short-term stock prices, trading volume, and investor behaviour [19]. Therefore, automated sentiment analysis can help users quickly understand the potential market implications of news events, reducing the reliance on manual reading and subjective judgment [20].

Compared with general-domain sentiment analysis, financial news sentiment classification is more challenging. Financial texts often contain domain-specific terminology, numerical indicators, and fixed expressions that cannot be accurately interpreted by general sentiment dictionaries. In addition, financial news usually contains a large proportion of objective and neutral statements, making fine-grained classification more difficult. The sentiment polarity of financial expressions is also highly context-dependent; for example, the same term may imply different market meanings under different corporate or macroeconomic conditions.

In investment analysis, news sentiment can serve as a useful complement to fundamental analysis and market data analysis. Prior studies have shown that financial news sentiment is associated with short-term stock return volatility, market liquidity, and investor trading behaviour [18]. For example, positive earnings-related news may strengthen investor confidence, while negative risk-related news may increase market concern and selling pressure. In this research, the proposed hybrid financial analysis agent uses a lightweight financial news sentiment classifier as a core middleware component. The classifier outputs standardized sentiment labels and sentiment intensity features, which are combined with RAG-retrieved financial knowledge, real-time market data, and corporate fundamental data to generate more comprehensive investment analysis feedback for end users [21].

2.5. Text Mining and Classification Methods in Financial Scenarios

This research adopts a traditional machine learning pipeline for financial news sentiment classification, including text preprocessing, TF-IDF feature extraction, supervised classifier training, and performance evaluation [22]. The design focuses on lightweight deployment, fast inference, and interpretability, which are important for integrating the sentiment module into a practical financial analysis agent.

2.5.1. Domain Text Feature Extraction Technology

TF-IDF is a widely used statistical text representation method that assigns weights to terms based on their frequency in a document and their inverse frequency across the corpus [23]. Compared with the basic Bag-of-Words model, TF-IDF reduces the influence of common words and gives higher importance to terms that are more discriminative for classification. In this research, TF-IDF with unigram and bigram features is used to represent financial news sentences. Unigram features capture individual financial keywords, while bigram features preserve short phrase patterns such as "price target" and "earnings forecast", which are useful for financial sentiment classification.

2.5.2. Supervised Text Classification Machine Learning Models

LinearSVC is selected as the main supervised classification model in this research. It is suitable for high-dimensional sparse text features generated by TF-IDF and provides efficient training and inference performance [22]. Compared with nonlinear SVM variants, LinearSVC avoids expensive kernel computation and is more practical for lightweight financial text classification tasks. In addition, the model uses class_weight='balanced' to reduce the influence of class imbalance among positive, neutral, and negative samples in the financial news corpus [24].

Transformer-based financial sentiment models, such as FinBERT, can capture richer contextual cues than sparse TF-IDF features [25]. However, they also require larger model checkpoints, higher inference latency, and greater runtime memory. Since the sentiment service is designed to run inside a real-time FastAPI agent workflow, this project treats FinBERT as a comparison baseline and uses TF-IDF with LinearSVC as the default deployment option because of its more favorable speed, memory, and deployment cost.

2.5.3. Standard Financial Sentiment Annotation Dataset

The sentiment classifier is trained and evaluated on the FinancialPhraseBank dataset proposed by Malo et al. [21]. The dataset contains 4,846 financial news sentences collected from financial media and company announcements, with sentiment labels assigned by annotators with financial background. It provides several subsets based on different levels of annotator agreement. This project uses the AllAgree subset, which contains samples where all annotators agreed on the same sentiment label, in order to reduce annotation noise and improve the consistency of model training and evaluation.

2.6. Research Gap and Theoretical Justification

Existing financial decision-support systems still have several limitations. Many traditional financial QA systems mainly rely on structured market data and company fundamentals, while paying less attention to unstructured news sentiment. As a result, they may fail to capture market reactions reflected in recent financial news. Meanwhile, many sentiment analysis studies focus on improving standalone classification accuracy, but do not integrate sentiment signals with RAG, real-time market data, fundamental data, and agent-based reasoning in a unified workflow. In addition, general-purpose large language models often lack access to up-to-date financial data and may produce unsupported or inaccurate answers in domain-specific scenarios. To address these gaps, this project develops a hybrid financial analysis agent that integrates RAG-based

knowledge retrieval, financial text mining, sentiment classification, and multi-source financial data access.

3. Data Science Methods

3.1. Overall System Architecture

The proposed system was designed as an integrated financial analysis agent that combines domain knowledge retrieval, real-time market data access, fundamental data access, news understanding, and answer generation in a unified workflow [26]. The architecture follows a modular full-stack design, with a React frontend for user interaction and a FastAPI backend for the core intelligent agent, as shown in Figure 1. The backend workflow is implemented with LangGraph, which represents the agent process as graph nodes for planning, tool execution, evidence accumulation, and final answer generation.

To avoid duplicating the diagram labels in the running text, this section summarizes only the functional role of each layer. The frontend collects natural-language questions, optional selected-news context, and conversation history. The FastAPI backend converts these inputs into an agent state. The LangGraph workflow then routes the state to the appropriate module, including RAG retrieval, market data access, fundamental data access, web/news evidence, or sentiment analysis. After the required evidence has been collected, the system generates a structured response and returns it to the frontend for display. This layer separation improves maintainability and extensibility, allowing new tools, data sources, or models to be added without redesigning the whole system [27].

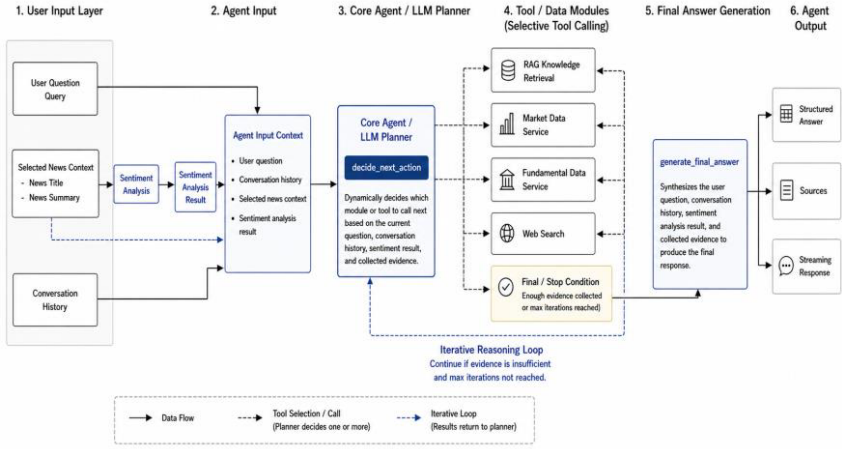


Figure 1. Agent System Architecture

3.2. Data Sources

This project used both static and dynamic data sources to support different financial analysis tasks. First, the RAG module used a local finance knowledge base consisting of a glossary dataset and an FAQ dataset. This domain-specific knowledge base design ensures accurate retrieval of professional financial concepts and avoids factual hallucinations common in general large language models [28]. The glossary file

contained financial and economics terms, definitions, topics, categories, and source information. The FAQ file contained man-made natural-language questions and their corresponding financial concept answers. These two datasets were used to support concept explanation and financial knowledge retrieval.

Second, the sentiment analysis component used the Financial PhraseBank dataset. This dataset contains labeled financial news sentences with three sentiment classes: positive, neutral, and negative. We use the AllAgree subset to train and evaluate the financial news sentiment classifier in this project. The distribution is shown in Figure 6, implying that the dataset is imbalanced, with neutral samples forming the largest class, followed by positive samples, while negative samples are the least represented.

Third, the system used live external APIs to obtain current financial information. Real-time API integration is a critical component of modern financial analysis systems, as it provides up-to-date market information that static datasets cannot offer [29]. The Twelve Data API was used to retrieve market data and company fundamental data, such as stock prices, returns, volatility, P/E ratio, EPS, and market capitalization. In addition, the Alpha Vantage news API was used to collect recent financial news, including titles, summaries, sources, and publication time. These live sources allowed the system to provide more up-to-date financial analysis.

Overall, the local datasets supported knowledge retrieval and sentiment modeling, while the external APIs supported real-time market and company analysis.

3.3. Data Cleansing and Preprocessing

Different preprocessing strategies were applied to different data sources in this project. Proper data preprocessing is important for improving NLP model performance and ensuring the reliability of downstream analysis tasks [30].

For the RAG knowledge base, the glossary and FAQ datasets were cleaned by removing records with missing values, duplicated entries, and noisy text. Their fields were then standardized into consistent structures. After cleaning, each glossary item was converted into a "term: definition" format, while each FAQ item was converted into a "question-answer" format for retrieval.

For the sentiment analysis dataset, the Financial PhraseBank samples were cleaned and standardized before model training. Invalid records were removed, and the original sentiment labels were retained as positive, neutral, and negative. The detailed feature extraction and classifier training process is described in Section 3.5.

For live API data, preprocessing mainly focused on format normalization, field validation, caching, and error handling. News data were standardized into fields such as title, summary, source, and timestamp. Market and fundamental data were transformed into structured indicators that could be directly used by the agent during answer generation.

3.4. RAG Retrieval Module Implementation

The RAG retrieval module is responsible for financial knowledge retrieval and concept-level question answering in the proposed system. It retrieves relevant financial knowledge chunks from the local knowledge base and provides them as evidence for downstream answer generation. The module uses the cleaned glossary and FAQ data described in Section 3.2 as its local knowledge base.

The retrieval workflow contains two parallel branches. In the dense retrieval branch, the user query is converted into an embedding using text-embedding-3-small, and semantic retrieval is performed through Redis Vector Search. In the sparse retrieval branch, keywords are extracted from the query and used for BM25-based or Redis keyword-based matching. The results from both branches are then merged using Reciprocal Rank Fusion (RRF) to produce a unified candidate list. For a document d returned by retrieval source r, RRF assigns the following score:

$$\mathrm{RRF}(d) = \sum_{r \in R} \frac{1}{k + \mathrm{rank}_r(d)}$$

where R denotes the set of retrieval result lists, $\mathrm{rank}_r(d)$ is the rank position of document d in list r, and k is a smoothing constant that reduces the dominance of top-ranked items from a single retriever [31]. After fusion, a reranker reorders the candidate chunks according to their relevance to the original query and their financial terminology match. Finally, the module outputs the Top-K financial chunks as the retrieved context for answer generation.

The implementation of the RAG module is mainly located in the backend/rag/service/ directory. The file rag_service.py handles knowledge base initialization, text chunk construction, embedding generation, and retrieval interface encapsulation. The file vector_store_service.py implements the hybrid retrieval mechanism, including Redis vector search and keyword-based retrieval. The file reranker_service.py provides a lightweight rule-based second-stage reranker. Instead of using a transformer cross-encoder, this reranker prioritizes fused retrieval rank, query-term overlap, and matches with finance-specific terminology, which keeps the retrieval pipeline reproducible and inexpensive for the prototype deployment.

The integration between the RAG module and the agent workflow is implemented in backend/agent/nodes.py and backend/agent/graph.py. When the agent determines that a user question requires financial knowledge support, the retrieve_knowledge_context node calls the RAG service and stores the retrieved chunks in the agent state. The final answer is generated by generate_final_answer, which combines RAG evidence with market data and other contextual information. The response is then wrapped by backend/api/runner.py and returned to the frontend.

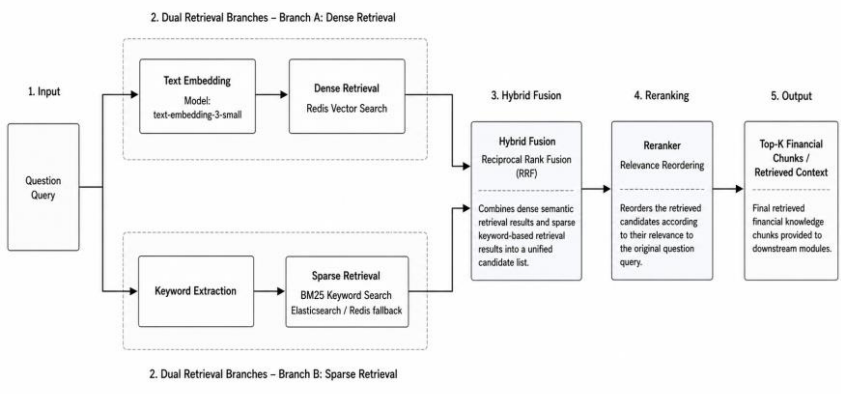


Figure 2. Hybrid RAG Retrieval Workflow

3.5. Sentiment Classification Model

The sentiment classification module uses a lightweight traditional machine learning pipeline to classify financial news into positive, neutral, and negative labels. This design was selected because it provides fast inference, low computational cost, and clear interpretability, which are suitable for integration into the hybrid financial analysis agent.

The pipeline includes text preprocessing, TF-IDF feature extraction, LinearSVC training, model evaluation, and backend deployment. In the preprocessing stage, noisy characters are removed, whitespace and letter case are normalized, and the text is converted into a clean input representation. The TF-IDF vectorizer then extracts unigram and bigram features with ngram_range = (1, 2) and max_features = 5000. The unigram and bigram setting allows the model to capture both individual financial terms and short phrase patterns, while the feature limit keeps the representation compact. LinearSVC is used as the core classification model because it is effective for high-dimensional sparse text features generated by TF-IDF. The main hyperparameters are set as C = 1 and class_weight = 'balanced'. The regularization parameter C = 1 provides a standard balance between fitting and generalization, while class_weight = 'balanced' helps reduce the influence of class imbalance among positive, neutral, and negative samples.

After offline training and evaluation, the final vectorizer and LinearSVC model are serialized and loaded by the FastAPI sentiment service. During runtime, the service receives the selected news title and summary, predicts a sentiment label, and adds the result to the LangGraph agent state. This design makes the connection between offline model training and live backend routing explicit: the trained classifier is not a separate experiment, but a deployed middleware component that contributes a sentiment signal to the broader financial analysis response.

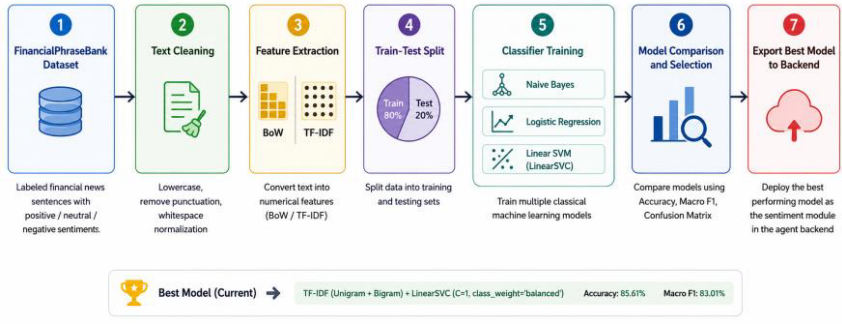


Figure 3. Sentiment Classification Model Training Pipeline

3.6. Agent Workflow and Tool Routing

The core logic of the system is implemented as a multi-step agent workflow. In this project, the agent was developed using a LangGraph architecture, a graph-based orchestration design, while the action planner is powered by GPT-5.4-mini. This allows different tools to be selected according to the information need of the question.

The workflow begins with an action planning stage. Once the user question is received, the system first examines the question itself, optional news context selected from the frontend, prior conversation history, and any evidence that has already been collected. Based on these inputs, the planner determines the next action from a predefined

set of tools: rag, market_data, fundamental, web_search, or finalize. This design allows the system to move iteratively through several reasoning steps rather than generating an answer immediately from the raw question.

If the planner decides that the user is asking for a definition, explanation, or concept-level knowledge, the query is routed to the RAG module. If the question is primarily about stock price movement, return, volatility, or recent trend, the system calls the market data service. If the question concerns valuation or company-level financial indicators, such as P/E ratio, EPS, revenue, or market capitalization, the system uses the fundamental data service. If the internal modules do not provide enough evidence, the agent can also use web search as a supplementary source. After one or more tools have been executed, the workflow returns to the planning stage again, where the system evaluates whether the current evidence is sufficient or whether another tool should be called.

This iterative strategy improves answer quality because it avoids relying on a single source of evidence. In addition, the system records tool usage history, intermediate outputs, and the current evidence snapshot. These records help the final generation stage produce answers that are more grounded and traceable. The workflow also includes a maximum iteration limit to prevent excessive tool calls and to ensure that the system eventually finalizes an answer within a reasonable number of steps.

Another practical design feature is the fallback logic. If a tool fails, for example due to missing ticker extraction, unavailable API data, or network problems, the agent does not terminate immediately. Instead, it stores the error state and attempts to continue reasoning with the remaining evidence sources. This makes the workflow more robust in realistic conditions where external services may be unstable or incomplete. Therefore, the agent workflow is not simply a single question-answer pipeline, but a controlled reasoning process that incrementally gathers evidence and synthesizes the final response.

3.7. Frontend Interaction and System Integration

The frontend layer of the proposed hybrid financial analysis agent is implemented with React and serves as the main interaction interface between users and the backend financial intelligence modules. The interface adopts a two-panel design: the left panel presents a market news feed, while the right panel provides an AI agent workspace for contextual question answering. This layout allows users to first browse financial news and then ask follow-up questions about the selected market event.

A key interaction mechanism is selected-news contextualization. When a user selects a news item, the frontend converts it into a structured NewsContext object rather than simply appending the news text to the query. This object contains the news title, summary, source, publication time, related ticker, and URL. By sending structured context to the backend, the system enables the agent to combine the user question with stock-specific metadata and route the task to suitable modules, such as sentiment analysis, market data retrieval, fundamental data retrieval, RAG, or web search. If no question is entered after a news item is selected, the frontend uses a default analysis prompt to reduce the prompt-writing burden for ordinary users.

The system mainly relies on streaming interaction to improve responsiveness and transparency. The frontend sends the user question, optional news context, and conversation history to the /api/stream endpoint. The backend then returns streaming

agent events, allowing the frontend to update the answer and workflow trace in real time. The trace is generated from the actual backend execution path, so different questions may display different tool selection and reasoning processes. For RAG-related queries, the trace can also show retrieved chunk summaries and source information, which helps users understand the evidence behind the generated answer.

The backend is implemented with FastAPI and exposes four main endpoints: /api/health, /api/news, /api/ask, and /api/stream. Among them, /api/news retrieves financial news, /api/ask supports standard request-response question answering, and /api/stream serves as the main endpoint for real-time frontend interaction. Overall, this integration design connects news browsing, structured context passing, streaming responses, and workflow trace visualization, supporting the project's goal of building an integrated financial decision-support assistant rather than a standalone chatbot.

4. Results and Interpretation

4.1. Sentiment Analysis Experimental Setup

The sentiment experiment evaluates which text classification setting is most suitable for the deployed news-understanding component of the hybrid financial analysis agent. The goal is not only to maximize offline accuracy, but also to identify a model that can be integrated into the live FastAPI backend with low inference latency, low memory use, and stable behavior under class imbalance.

All experiments use the AllAgree subset of the FinancialPhraseBank dataset. The labelled samples are split into training and test sets using an 8:2 ratio with a fixed random seed, so that results remain reproducible across model configurations. Since the subset contains many neutral samples and fewer directional positive or negative samples, class imbalance is treated as an important part of the evaluation setting.

The evaluation reports Accuracy, Macro F1, Weighted F1, per-class F1, and the confusion matrix. Macro F1 is emphasized because it gives equal weight to negative, neutral, and positive classes, while the per-class F1-scores and confusion matrix show whether the model produces asymmetric errors such as false-positive sentiment flags. The experiment compares Bag-of-Words and TF-IDF features, Naive Bayes, Logistic Regression, and LinearSVC classifiers, different LinearSVC regularization settings, and a FinBERT inference baseline for deployment-cost comparison.

4.2. Sentiment Model Performance

The sentiment classification experiment shows in Table 1 that TF-IDF-based features worked effectively for financial news sentiment classification. Among the tested settings, the best-performing configuration used the AllAgree dataset, TF-IDF unigram and bigram features, and a balanced LinearSVC classifier with C = 1. This model achieved a Macro F1 score of 83.01% and an Accuracy of 87.61%.

Table 1. Results of Tfidf.

Rank	Dataset	Features	Classifier	Macro F1	Accuracy
1	AllAgree	TF-IDF uni+bi, loose	LinearSVC balanced, C = 1	83.01%	87.61%
2	AllAgree	TF-IDF / LinearSVC setting	LinearSVC balanced, C = 2	82.24%	87.17%

Rank	Dataset	Features	Classifier	Macro F1	Accuracy
3	AllAgree	TF-IDF uni+bi, loose	LinearSVC, C = 1	82.15%	87.39%
4	AllAgree	TF-IDF uni+bi, loose	LinearSVC balanced, C = 5	81.87%	87.17%
5	AllAgree	TF-IDF uni+bi, loose	Logistic Regression balanced, C = 5	81.73%	86.50%

To address the class-imbalance issue more directly, Table 2 reports the per-class F1-scores and the deployment-cost comparison between the selected TF-IDF + LinearSVC classifier and a FinBERT baseline.

Table 2. Results of TF-IDF + LinearSVC classifier and a FinBERT baseline.

Model	Acc.	Macro F1	Weighted F1	Neg. F1	Neu. F1	Pos. F1
TF-IDF + LinearSVC	0.8761	0.8301	0.8763	0.7642	0.9367	0.7895
FinBERT	0.9690	0.9593	0.9694	0.9365	0.9797	0.9617

Model	Total time	Avg./sample	Memory increase	Model size
TF-IDF + LinearSVC	0.0054 s	0.0120 ms	0.0000 MB	0.9755 MB
FinBERT	12.5543 s	27.7750 ms	28.2344 MB	417.9261 MB

The TF-IDF + LinearSVC classifier achieves negative, neutral, and positive F1-scores of 0.7642, 0.9367, and 0.7895 respectively. These values show that the model is strongest on the neutral class and weaker on the two directional classes. The confusion matrix further indicates a limited false-positive pattern: 12 truly negative samples and 12 truly neutral samples are predicted as positive. This matters in a financial agent because positive false positives may overstate market optimism. Therefore, the final interface should treat the sentiment label as an auxiliary signal rather than as a standalone investment conclusion.

FinBERT achieves stronger classification performance, with 0.9690 accuracy and 0.9593 Macro F1. However, its computational cost is substantially higher for the deployment scenario. In the same test run, TF-IDF + LinearSVC requires 0.0120 ms per sample and a model size of about 0.9755 MB, while FinBERT requires 27.7750 ms per sample and about 417.9261 MB of model storage. For this reason, FinBERT is useful as an accuracy-oriented baseline, but TF-IDF + LinearSVC remains the default deployed classifier because it offers a better cost-performance trade-off for a real-time FastAPI agent.

The results suggest three main findings. First, TF-IDF features performed better than simpler text representation settings. This is reasonable because financial news often depends on specific terms and short phrases, such as “earnings growth,” “market uncertainty,” “profit warning,” or “rate cut expectation.” TF-IDF can give higher importance to discriminative words and phrases, which makes it suitable for financial sentiment classification.

Second, LinearSVC achieved the strongest overall balance among lightweight classical models. Compared with Logistic Regression, LinearSVC produced better Macro F1 and Accuracy in the reported top settings. This indicates that a linear margin-based classifier is effective for sparse high-dimensional financial text features.

Third, the AllAgree dataset produced more reliable results. Since the labels in this subset have higher annotator agreement, the training signal is cleaner. This helps the model learn clearer boundaries between positive, neutral, and negative sentiment.

4.3. Integrated System Demo Results

The integrated system demo shows that the project is not only a standalone sentiment classifier, but a broader RAG-based financial QA system. The agent takes the user query, selected news context, and conversation history as input. It then uses a LangGraph-based workflow to decide which module should be called next. Depending on the query type, the system may use the RAG module, market data module, fundamental data module, web/news module, and sentiment classifier before generating the final answer.

The demo results show that the system can support several common financial analysis tasks. For general finance-related questions, the RAG module retrieves relevant financial knowledge and returns context for the final answer. For company-specific questions, the market data and fundamental modules provide numerical information such as price movement, return, volatility, trend, P/E ratio, EPS, market capitalization, and other statistics. For news-related questions, the news API provides recent news title, summary, source, and publication time. The sentiment classifier then adds an additional signal by identifying whether the selected news is positive, neutral, or negative.

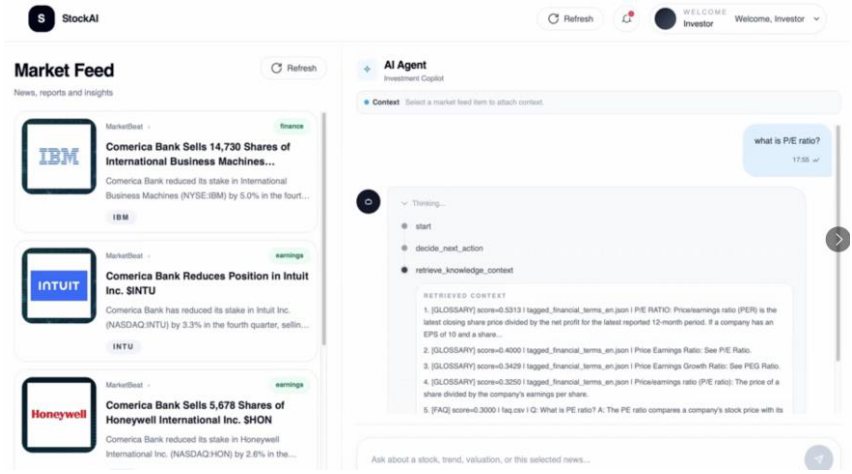


Figure 4. Example Thinking when user type in a question

4.4. Interpretation of Findings

The results show that the project successfully demonstrates the value of combining RAG, text mining, and external financial data APIs in one agent system. The sentiment model provides a measurable machine learning component, while the integrated agent workflow shows how this model can be used in a practical financial analysis pipeline.

The sentiment classification result is adequate for a prototype-level financial analysis agent. An accuracy of 87.61% and Macro F1 of 83.01% suggest that the lightweight model can distinguish financial sentiment reasonably well under the selected dataset setting. More importantly, the per-class F1-scores show that the evaluation does not only reward performance on the majority class. The lower F1-scores for negative and

positive samples also clarify the model's limitations and justify using sentiment as a supporting signal rather than a direct recommendation.

The integrated demo also shows that the system is more useful than a simple chatbot. A simple chatbot usually produces a direct natural language answer, but it may not show where the answer comes from or whether it is supported by updated data. The proposed system improves this by separating the workflow into multiple modules. The RAG module supports knowledge-based answers, the market data module supports price and trend analysis, the fundamental module supports company-level financial analysis, and the sentiment classifier supports news interpretation.

However, the results should be interpreted as a prototype-level evaluation rather than a full financial decision-making benchmark. The current system demonstrates technical feasibility, but it does not guarantee investment accuracy. Financial markets are affected by many unpredictable factors, and sentiment is only one signal among many. Therefore, the system should be understood as a decision-support assistant, not as an automated trading or investment recommendation system.

4.5. Practical Value for Users

The practical value of the system is that it reduces the effort required for ordinary users to understand financial information. Instead of checking news websites, market data platforms, financial statements, and financial terminology separately, the user can ask one question and receive a structured answer.

The system is especially useful in three situations. First, it helps users quickly understand the meaning of financial news. Second, it helps users connect news sentiment with market and fundamental data. Third, it helps users ask follow-up questions in a conversational way, which makes the analysis process more accessible.

Overall, the results support the main aim of the project: to build an integrated decision-support assistant for financial information analysis. The current prototype shows the core workflow, including retrieval, tool routing, external data access, sentiment classification, and structured answer generation. These results should be interpreted as evidence of improved information organization and access for users, rather than proof that the system directly improves market efficiency. Future improvements can further increase the system's reliability by expanding the financial knowledge base, adding more financial indicators, improving risk warning signals, and making the output more transparent.

5. Discussion

5.1. How the system addresses the project goals

The primary goal of this project was to mitigate information overload and simplify financial research for everyday investors. The developed system successfully addresses this by deploying an autonomous Large Language Model (LLM) agent that acts as a centralized intelligence hub. Using a question-based workflow, the LLM planner dynamically routes user queries to the appropriate analytical tools—including a RAG module, real-time market data APIs, fundamental data APIs, web search, and a sentiment analysis module. This architecture actively bridges isolated data points, translating

fragmented financial signals into coherent, actionable insights and significantly reducing the user's cognitive load.

5.2. Strengths of the proposed solution

The core strength of this system lies in its hybrid agent architecture, which extends a pure RAG implementation with live data access and a deployed text mining component. While a pure RAG pipeline is limited to retrieving static, pre-indexed text, the proposed system integrates live external APIs (Twelve Data and Alpha Vantage) to fetch current market trends and fundamental metrics. Furthermore, the system is augmented by a dedicated sentiment classifier trained on the FinancialPhraseBank dataset using TF-IDF and LinearSVC. Achieving 87.61% accuracy and 83.01% Macro F1, this classifier provides a lightweight sentiment signal that can be combined with RRF-based retrieval, market data, and company fundamentals for multi-source analysis.

5.3. Limitations of the current implementation

Despite its strengths, the current implementation has several notable limitations. First, RAG quality remains an ongoing challenge; evidence ranking and source filtering need further optimization to prevent noisy or irrelevant textual context from influencing the LLM's final output. Second, metric coverage is somewhat restricted. The system currently evaluates a limited set of financial indicators (such as P/E and EPS) and lacks integration with broader macroeconomic data or advanced technical analysis metrics. Finally, risk and safety mechanisms require substantial improvement. Because financial decisions carry real-world consequences, the system needs stronger built-in risk explanations and higher transparency to ensure users do not blindly rely on AI-generated outputs without human verification.

5.4. Comparison with simpler chatbot approaches

Simpler, standard chatbots rely mainly on parametric memory and user-provided context. This makes them vulnerable to outdated or unsupported answers in fast-moving financial settings, especially when current stock prices, company indicators, or breaking news are required. In contrast, the proposed agent workflow enforces a structured evidence-gathering process. By selecting external tools to fetch live structured data, extracting current fundamental metrics, and classifying selected news sentiment before generating a response, the system grounds its answers in explicit evidence sources and reduces reliance on unsupported generation.

5.5. Implications for financial information analysis

This project illustrates the practical value of combining autonomous LLM agents with classical machine learning classifiers and specialized financial APIs. By automating the extraction, classification, and synthesis of heterogeneous financial information, the system lowers the effort required for users to interpret market data, company fundamentals, and news context. Its main implication is therefore user-level information support: it helps non-expert investors organize relevant evidence more efficiently, while still requiring human judgment for final financial decisions.

6. Conclusions

6.1. Conclusion

This project developed an intelligent financial assistant designed to mitigate information overload for everyday investors. By integrating a LangGraph-based agent with a hybrid RAG pipeline, real-time market APIs, and a custom news sentiment classifier, the system translates fragmented data into coherent informational responses. Compared with a standard chatbot, the system provides stronger evidence grounding through structured tool use and external data access. However, the current results should be interpreted as prototype-level evidence. Limitations remain regarding RAG evidence ranking, restricted metric coverage, sentiment false positives, and the need for greater risk transparency.

6.2. Future work

To address current constraints and enhance system reliability, future development will focus on three key areas:

- Knowledge & Metrics: Expanding the financial knowledge base and integrating a broader range of macroeconomic and technical indicators to improve analytical depth.
- Explainability & RAG Quality: Upgrading evidence ranking and source filtering within the RAG pipeline. The system will be enhanced to display precise reasoning steps, evidence quality, and explicit sources to users.
- Risk & Safety Signals: Implementing confidence indicators and structured risk alerts to improve safety and transparency, ensuring users are fully aware of the inherent risks associated with financial data interpretations.

References

[1] Bernales A, Valenzuela M, & Zer I. Effects of information overload on financial market returns: How much is too much? SSRN Electronic Journal. 2021;doi.org/10.2139/ssrn.3904916.

[2] Tairu Bello S. The fundamentals of stock market investing in the twenty-first century. In Financial literacy in today's global market. IntechOpen. 2024;doi.org/10.5772/intechopen.1003853.

[3] Wang Z. Investor sentiment and the stock market. Highlights in Business, Economics and Management. 2023;21:346–351, doi.org/10.54097/hbem.v21i.14424.

[4] Al-Fattal A. When More Is Less: Information Overload and the Psychology of Decision-Making in Cryptocurrency Investment. Psychology International. 2026;8(1):17, doi.org/10.3390/psycholint8010017.

[5] Bernales A, Valenzuela, M., & Zer, I. Effects of information overload on financial markets: How much is too much? European Financial Management Association (EFMA) Annual Meetings; c2024. http://www.efmaefm.org/0EFMAMEETINGS/EFMA%20ANNUAL%20MEETINGS/2024-Lisbon/papers/paper_news2.pdf.

[6] Rossi AG, Utkus SP. Set it and forget it: Engineering investment habits with FinTech. Rutgers Business School Working Paper;c2024. https://www.business.rutgers.edu/sites/default/files/documents/set-it-and-forget-it.pdf.

[7] Lewis P, Perez E, Piktus A, Petroni F, Karpukhin V, Goyal N, Küttler H, Lewis M, Yih W, Rocktäschel, T, Riedel S, Kiela D. Retrieval-augmented generation for knowledge-intensive NLP tasks. Advances in Neural Information Processing Systems. 2020; 33:9459–9474.

[8] Nie Y, Kong Y, Dong X, Mulvey J M, Poor HV, Wen Q, & Zohren S. A Survey of Large Language Models for Financial Applications: Progress, Prospects and Challenges; c2024. arXiv. https://doi.org/10.48550/arxiv.2406.11903.

[9] Ji R, Han Q. Understanding heterogeneity of investor sentiment on social media: A structural topic modeling approach. Frontiers in Artificial Intelligence. 2022;5, doi.org/10.3389/frai.2022.884699.

[10] Cheung L, Wang Y, Lau ASM, Chan RMC. Using a Novel Clustered 3D-CNN Model for Improving Crop Future Price Prediction. Knowledge-based System. 2023. https://doi.org/10.1016/j.knosys.2022.110133.

[11] Nie RX, Liu ZY, Yu WY, Zhan RQ, Lau ASM. The Application of Artificial Intelligence in Financial Prediction. Frontier in Artificial Intelligence and Applications. 2025;412, https://doi.org/10.3233/FAIA250732.

[12] Karpukhin V, Oğuz B, Min S, Lewis P, Wu L, Edunov S, Chen D, Yih W. Dense passage retrieval for open-domain question answering. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP); Association for Computational Linguistics; c2020. p. 6769–6781, doi.org/10.18653/v1/2020.emnlp-main.550.

[13] Thakur N, Reimers N, Rücklé A, Srivastava A, Gurevych I. BEIR: A heterogeneous benchmark for zero-shot evaluation of information retrieval models. Advances in Neural Information Processing Systems. 2021;34.

[14] Li H, Lu W, Wang J, Liu M. A survey of neural retrieval and reranking for question answering. Information Processing & Management. 2021;58(6):102701.

[15] Gao Y, Xiong Y, Gao X, Jia K, Pan J, Bi Y, Dai Y, Sun J, Wang M, Wang H. Retrieval-augmented generation for large language models: A survey; c2024. https://arxiv.org/abs/2312.10997.

[16] Yang P, Fang H, Lin J. Anserini: Enabling the use of Lucene for information retrieval research. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval; Association for Computing Machinery; c2017. p. 1253–1256, https://doi.org/10.1145/3077136.3080721.

[17] Nogueira R, Cho K. Passage re-ranking with BERT; c2019. arXiv. arXiv:1901.04085.

[18] Kearney C, Liu S. Textual sentiment in finance: A survey of methods and models. International Review of Financial Analysis. 2014;33:171–185,doi.org/10.1016/j.irfa.2014.02.006.

[19] Tetlock PC. Giving content to investor sentiment: The role of media in the stock market. The Journal of Finance. 2007;62(3):1139–1168, doi.org/10.1111/j.1540-6261.2007.01232.x.

[20] Loughran T, McDonald B. Textual analysis in accounting and finance: A survey. Journal of Accounting Research. 2016;54(4):1187–1230, doi.org/10.1111/1475-679X.12123.

[21] Malo P, Sinha A, Korhonen P, Wallenius J, Takala P. Good debt or bad debt: Detecting semantic orientations in economic texts. Journal of the Association for Information Science and Technology. 2014;65(4):782–796, doi.org/10.1002/asi.23062.

[22] Sebastiani F. Machine learning in automated text categorization. ACM Computing Surveys. 2002;34(1): 1–47, doi.org/10.1145/505282.505283.

[23] Salton G, Buckley C. Term-weighting approaches in automatic text retrieval. Information Processing & Management, 24(5). 1988:513–523, doi.org/10.1016/0306-4573(88)90021-0.

[24] Fan RE, Chang KW, Hsieh CJ, Wang XR, Lin CJ. LIBLINEAR: A library for large linear classification. Journal of Machine Learning Research. 2008;9:1871–1874.

[25] Araci D. FinBERT: Financial sentiment analysis with pre-trained language models; c2019. arXiv. https://doi.org/10.48550/arXiv.1908.10063.

[26] Li Y, Wang S, Ding H, Chen H. Large language models in finance: A survey. In Proceedings of the 4th ACM International Conference on AI in Finance. Association for Computing Machinery: c2023. doi.org/10.1145/3604237.3626869.

[27] Taivalsaari A, Mikkonen T, Pautasso C, Systä K. Full stack is not what it used to be. In M. Brambilla, R. Chbeir, F. Frasincar, & I. Manolescu (Eds.), Proceedings Web Engineering: 21st International Conference, ICWE 2021: Springer; c2021. p. 363–371, doi.org/10.1007/978-3-030-74296-6_28.

[28] Barron RC, Grantcharov V, Wanna S, Eren, ME, Bhattarai M, Solovyev N, Tompkins G, Nicholas C, Rasmussen KØ, Matuszek C, Alexandrov BI. Domain-specific retrieval-augmented generation using vector stores, knowledge graphs, and tensor factorization; c2024. arXiv. arXiv:2410.02721.

[29] Ekster G, Kolm PN. Alternative data and artificial intelligence in investment management. The Journal of Financial Data Science. 2021; 3(2):1–16.

[30] HaCohen-Kerner Y, Miller D, Yigal Y. The influence of preprocessing on text classification using a bag-of-words representation. PLOS ONE. 2020;15(5):e0232525, doi.org/10.1371/journal.pone.0232525.

[31] Cormack GV, Clarke CLA, Buettcher S. Reciprocal rank fusion outperforms Condorcet and individual rank learning methods. In Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval, Association for Computing Machinery; c2009. p.758–759. https://doi.org/10.1145/1571941.1572114.

A. Appendix

A.1. Results Summary

A.1.1. Prediction Results

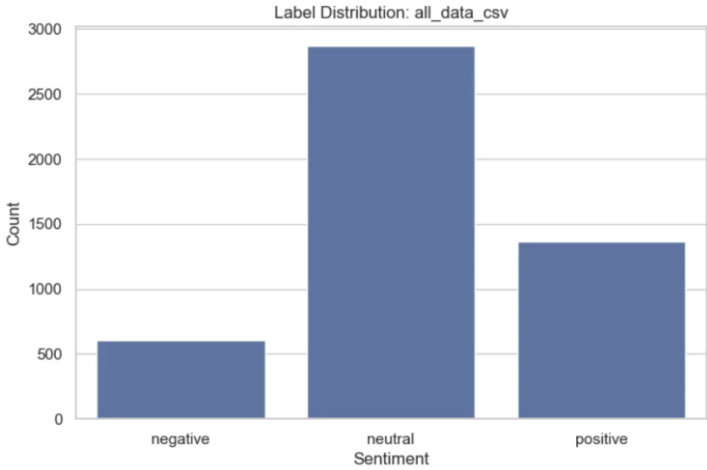


Figure 5. Sentiment Label Distribution in the raw FinancialPhraseBank Dataset

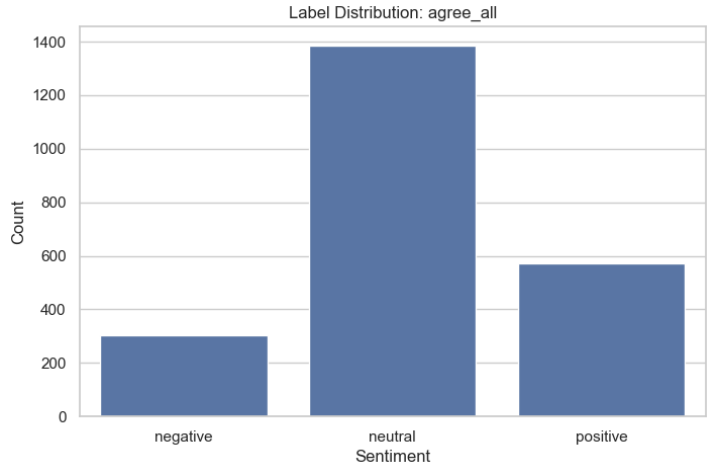


Figure 6. Sentiment Label Distribution in the AllAgree Subset of FinancialPhraseBank

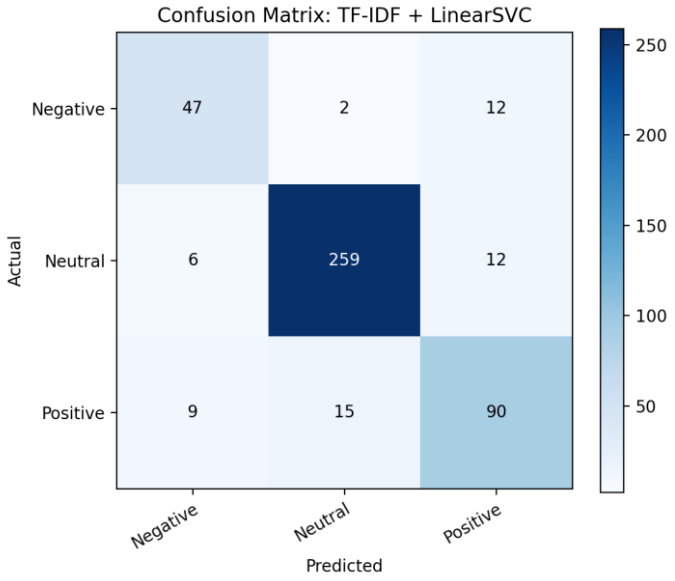


Figure 7. Confusion Matrix of the TF-IDF + LinearSVC Sentiment Classifier

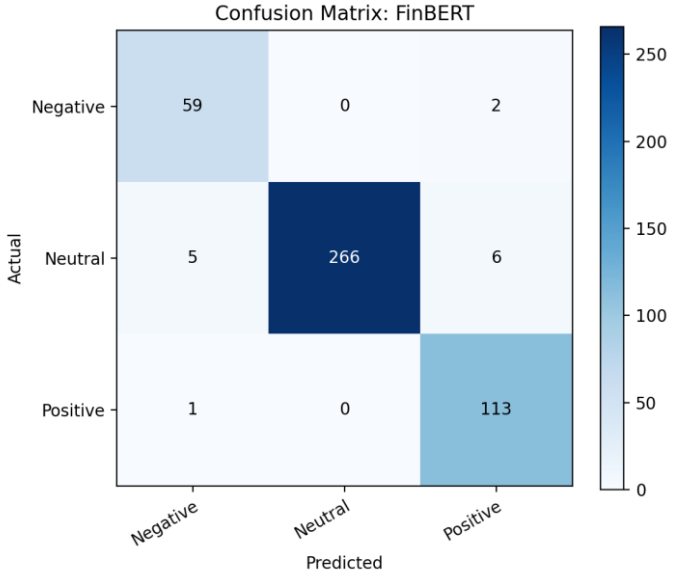


Figure 8. Confusion Matrix of the FinBERT Baseline

A.1.2. System Demonstration Record

Purpose: This appendix converts the recorded StockAI prototype demonstration into static report evidence. The figures document the main workflow stages: market-feed display, RAG-based financial QA, context-aware news analysis, market-data retrieval, and API error handling.

Evaluation role: The screenshots are used as qualitative system evidence for Chapter 4 Results and Interpretation and Appendix A Results Summary. They demonstrate implemented functionality rather than investment performance.

Table A.1. Demonstration sequence and corresponding system evidence.

Step	Demo stage	System component	Observed evidence	Figure
1	Interface overview	Frontend + market feed	Two-panel layout with market-news cards and AI Agent chat.	A.1
2	Concept query: "what is P/E ratio?"	RAG-based financial QA	Trace shows decide_next_action and retrieve_knowledge_context.	A.2
3	P/E answer generation	Answer generation	Defines P/E, explains interpretation, and gives an example.	A.3
4	IBM news selected for impact analysis	Context attachment + news workflow	News card is attached as context before the impact query.	A.4
5	IBM news-impact response	Sentiment-aware news analysis	Output labels the impact as mildly negative to neutral and explains risks/opportunities.	A.5
6	Tesla recent price trend query	Market data module	Reports latest close, seven-day return, daily move, and "sideways" trend label.	A.6

Note. The system output is for informational decision support only and should not be interpreted as investment advice.

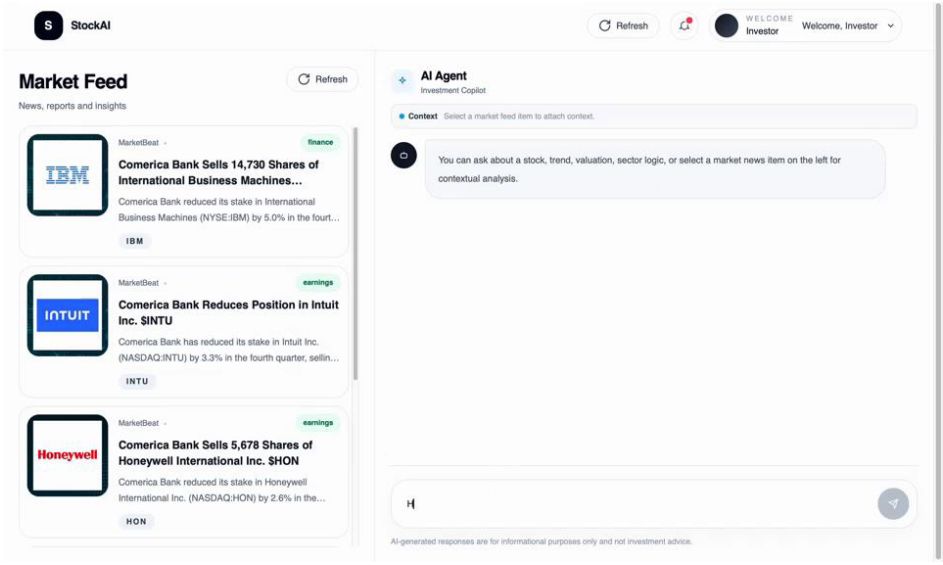


Figure A.1. Main interface of the StockAI financial analysis agent.

Explanation. The main interface separates market-feed browsing from conversational analysis. This design allows users to ask a general financial question or select a specific news item as context.

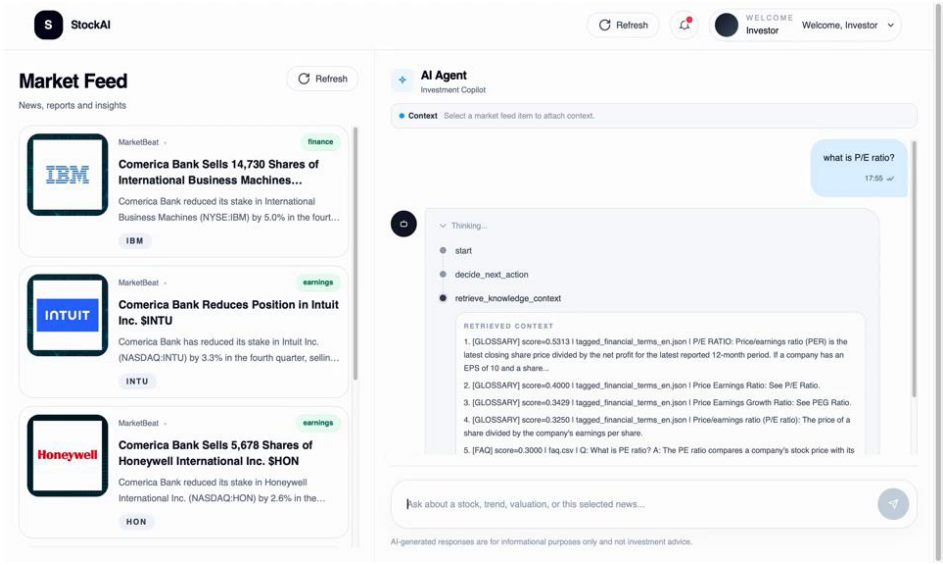


Figure A.2. RAG-based tool routing for a conceptual financial question on the P/E ratio.

Explanation. The agent displays intermediate workflow steps for the P/E-ratio query, including the retrieval of financial knowledge context before answer generation.

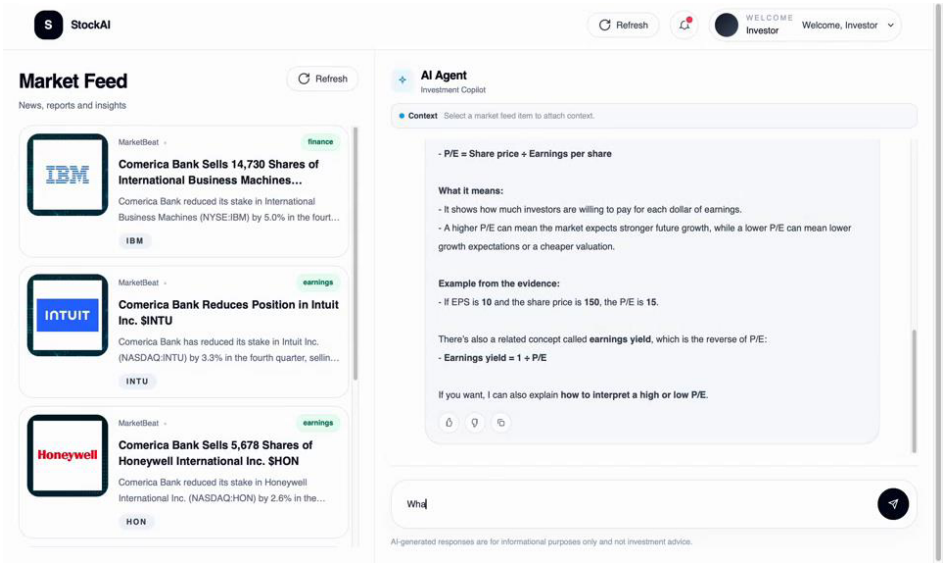


Figure A.3. Generated answer explaining the definition and interpretation of the P/E ratio.

Explanation. The final response provides a concise definition, interpretation logic, and example calculation, demonstrating the RAG-based financial QA function.

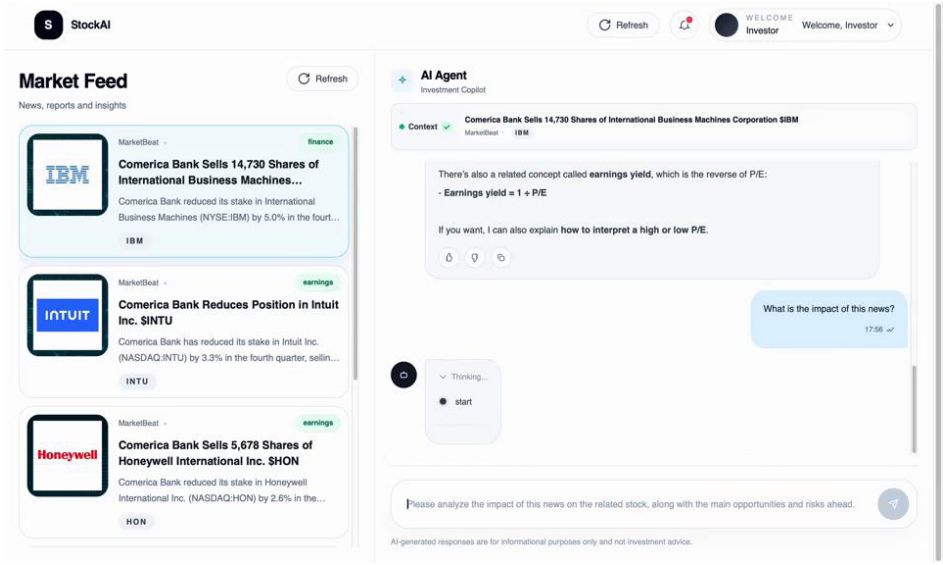


Figure A.4. Context-aware news selection and user query for analysing the impact of an IBM-related market news item.

Explanation. The selected IBM-related news item is attached as context. The user then asks the system to analyse the likely impact of the news on the related stock.

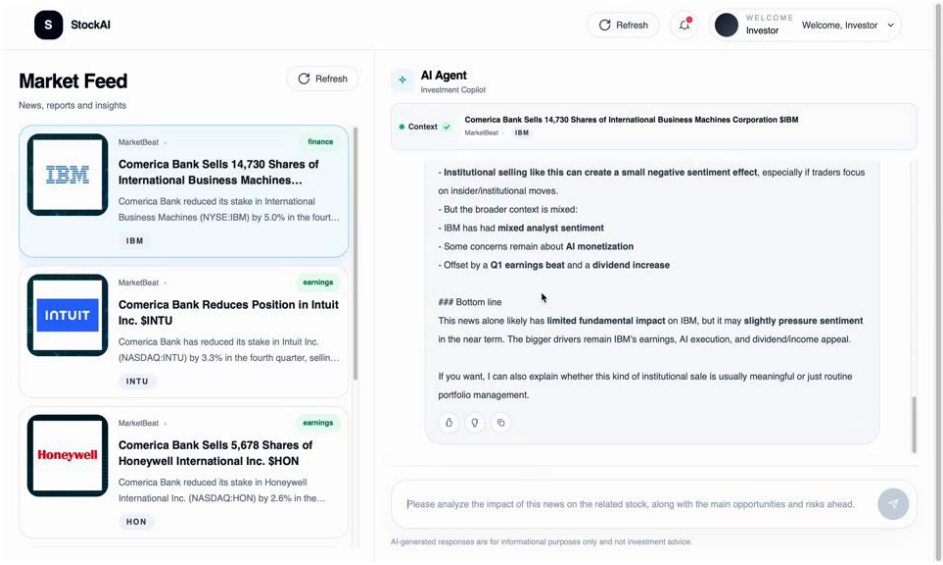


Figure A.5. News-impact analysis output, including direct sentiment interpretation and risk/opportunity discussion.

Explanation. The response interprets the selected news as mildly negative to neutral in the short term and explains why the signal

may be limited when considered with broader company-specific context.

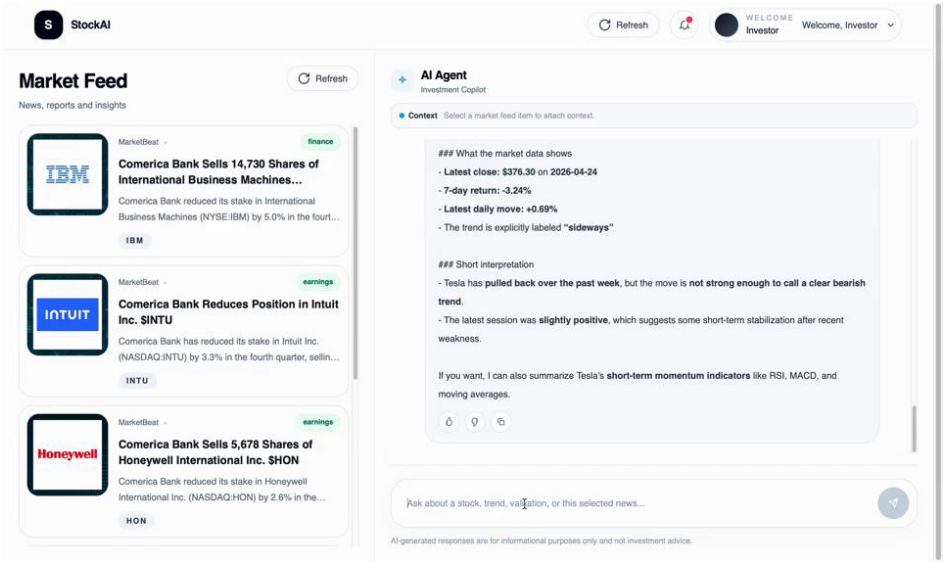


Figure A.6. Market-data response for Tesla, including latest close, seven-day return, daily movement, and trend label.

Explanation. The market-data module returns structured indicators for Tesla, including the latest close, seven-day return, latest daily move, and a sideways trend label.

A.2. Program Documentation

A.2.1. System Overview

The program implements a full-stack financial question-answering system that integrates a React frontend, FastAPI backend, LangGraph-based agent workflow, RAG-based financial knowledge retrieval, external financial data APIs, and a financial news sentiment classifier. The system is designed to help users understand stocks, market trends, financial concepts, company fundamentals, and the possible implications of financial news.

The overall workflow is as follows. First, the frontend loads recent financial news through the backend news endpoint. The user can either select a news item from the market feed or directly enter a general financial question. If a news item is selected, the frontend attaches the selected news context to the request. The frontend then sends the question, context, and recent conversation history to the backend. The backend constructs an agent state and runs the LangGraph workflow. Depending on the question and available evidence, the agent may call RAG retrieval, market data retrieval, fundamental data retrieval, web search, and sentiment analysis. The final answer is then streamed back to the frontend and displayed in the chat interface together with the execution trace.

A.2.2. Frontend Module

The frontend is organised around the main HomePage.tsx page and several reusable components. The main page coordinates the global interaction state, including news feed items, selected news context, chat messages, user input, loading state, error state, and active trace state.

The MarketFeedPanel component renders the left-hand news feed. It receives the list of feed items, the selected item ID, the refresh handler, and the item selection handler. When the user selects a feed item, the selected item becomes the active context for the next agent query.

The AgentPanel component renders the right-hand AI agent workspace. It includes the agent header, selected news context card, chat message area, error display, and chat input. The panel also provides different starter text depending on whether a news item has been selected. If a selected news item exists, the starter text encourages the user to ask about the impact of the news, opportunities, risks, valuation, and trend. If no item is selected, it encourages the user to ask general questions about stocks, trends, valuation, or sector logic.

A.2.3. Frontend API Client

The frontend communicates with the backend through the API client in frontend/src/services/api.ts. This file defines functions for fetching market news, sending one-off questions, and consuming streaming question-answering events.

The fetchMarketNews() function calls the /api/news endpoint with two query parameters: limit and force_refresh. The raw market news response is transformed into frontend feed items, including ID, title, summary, source, relative time label, category, ticker, image, and URL.

The streamQuestion() function is responsible for the main agent interaction. It constructs an AskRequest payload containing the user question, optional news context, and recent conversation history. It then posts the payload to /api/stream. The function reads the response body as a stream, decodes it as text, splits it by newline, parses each JSON line as a stream event, and passes each event to the frontend event handler.

A.2.4. Request and Response Schema

The frontend and backend share a consistent request-response structure. The main request object contains three fields:

Field	Description
question	The user's current financial question
context	Optional selected news context
history	Recent conversation history

The selected news context contains:

Field	Description
title	News title
summary	News summary
source	News source

Field	Description
time_ago	Relative publication time
ticker	Related stock ticker, if available
url	Original news URL

The backend response includes the following fields:

Table A.2. Backend response fields and descriptions.

Field	Description
request_id	Unique request identifier
question	Original user question
question_type	Inferred question type, such as asset or knowledge
classifier_result	Agent classification or planning result
asset_data	Market-related data if retrieved
rag_data	RAG retrieval result if used
sources	Evidence sources used by the answer
answer	Final generated response
error	Error message if execution fails

This schema makes the system easier to integrate and debug because the frontend does not receive only a final text answer. It also receives metadata about the agent's decision, retrieved evidence, and possible errors.

A.2.5. Backend API Module

The backend API is implemented in backend/api/main.py using FastAPI. It defines four main endpoints:

Table A.3. Backend API endpoints.

Method	Endpoint	Function
GET	/api/health	Returns backend health status
GET	/api/news	Retrieves recent financial news
POST	/api/ask	Runs a single non-streaming agent request
POST	/api/stream	Streams agent execution events through NDJSON

The /api/stream endpoint wraps the backend agent runner in a StreamingResponse and returns data with the media type application/x-ndjson. This design enables real-time frontend updates and supports the trace-based user experience.

A.2.6. Agent Runner

The backend runner constructs the initial state for each agent execution. The initial state includes the user question, optional news context, normalised conversation history, iteration count, maximum iteration limit, and tool history. The maximum iteration value is set to four, which prevents the agent from entering an uncontrolled loop.

In non-streaming mode, the runner invokes the compiled graph once and returns the final response. In streaming mode, the runner first emits a start event, then streams each LangGraph node update as a node event, and finally emits a final event containing the completed response. If an exception occurs, it emits an error event. This event-based design is directly connected to the frontend trace mechanism.

A.2.7. LangGraph Workflow

The agent workflow is defined in backend/agent/graph.py. The graph contains six main nodes:

Table A.4. LangGraph workflow nodes.

Node	Function
decide_next_action	Plans the next tool or decides to finalise
retrieve_knowledge_context	Retrieves relevant financial knowledge from the RAG system
get_market_data	Retrieves stock price and trend data
get_fundamental_data	Retrieves company valuation and financial metrics
execute_web_search	Searches for external web information
generate_final_answer	Synthesises evidence into the final answer

The workflow begins at decide_next_action. If the agent chooses RAG, market data, fundamental data, or web search, the graph routes the state to the corresponding tool node. After the tool node finishes, the graph returns to decide_next_action. This loop continues until the agent selects finalize, after which the graph moves to generate_final_answer and ends.

A.2.8. Sentiment and Evidence Integration

When a selected news context is present, the decide_next_action node checks whether sentiment has already been computed. If not, it combines the news title and summary and passes the text to the sentiment prediction service. The resulting sentiment is then included in the planner payload as part of the available evidence. This means the sentiment classifier is not used as an isolated output. Instead, it becomes one signal inside the broader agent reasoning and answer generation process.

The final answer generation node receives the user question, news context, news sentiment, conversation history, tool history, and current evidence snapshot. It then asks the language model to generate a final response based on the collected evidence. If RAG data exists, the generated answer is also inserted into the rag_data object so that the response remains consistent across the frontend schema.

A.2.9. Trace and Explainability

The frontend trace system is implemented through traceUtils.ts. Each streamed event is converted into a trace step with a pending, active, or completed status. When the trace is finalised, all steps are marked as completed.

A particularly useful feature is the retrieval detail display for the retrieve_knowledge_context node. If the node output contains retrieved chunks, the trace utility extracts the first five chunks and displays their chunk type, relevance score, source file, and text preview. This improves transparency for RAG-based answers and helps users or developers inspect whether the retrieved context is relevant to the question.

A.3. Program Coding

The implementation is organised into frontend, backend API, agent workflow, RAG, external services, and model components.

Table A.5. Main program modules and source files.

Module	File / Folder	Description
Main frontend page	frontend/src/pages/HomePage.tsx	Manages feed state, selected news context, chat messages, input, loading, error, and trace state
Agent panel	frontend/src/components/agent/AgentPanel.tsx	Renders selected context, chat messages, error display, and chat input
Market feed panel	frontend/src/components/feed/MarketFeedPanel.tsx	Renders the market feed and supports item selection and refresh
Trace utilities	frontend/src/components/agent/tracing/traceUtils.ts	Builds inline workflow trace and retrieved context details
Frontend API client	frontend/src/services/api.ts	Handles /api/news, /api/ask, and /api/stream requests
Frontend types	frontend/src/types/index.ts	Defines TypeScript interfaces for requests, responses, feed items, stream events, and trace state
FastAPI entry	backend/api/main.py	Defines API endpoints and streaming response
Agent runner	backend/api/runner.py	Constructs initial state, invokes graph, and streams NDJSON events
Backend schemas	backend/api/schemas.py	Defines Pydantic request and response models
LangGraph workflow	backend/agent/graph.py	Defines graph nodes, conditional routing, and graph edges
Agent nodes	backend/agent/nodes.py	Implements planning, RAG retrieval, market data, fundamentals, web search, and final answer generation
Agent state	backend/agent/state.py	Defines the TypedDict structure used across graph nodes
RAG service	backend/rag/service/rag_service.py	Orchestrates retrieval and context construction
Market data service	backend/services/market_data_service.py	Retrieves market price and trend information
Fundamental data service	backend/services/fundamental_data_service.py	Retrieves company valuation and financial metrics
News API service	backend/services/news_api.py	Retrieves recent financial news
Sentiment service	backend/services/sentiment_service.py	Loads the trained TF-IDF and LinearSVC model for news sentiment prediction

A typical streaming request sent from the frontend to the backend follows this structure:

```
{
  "question": "Please analyze the impact of this news on the related stock, alon
g with the main opportunities and risks ahead.",
  "context": {
    "title": "Example financial news title",
    "summary": "Example news summary",
    "source": "Yahoo Finance",
    "time_ago": "2h ago",
    "ticker": "AAPL",
    "url": "https://example.com/news"
  },
  "history": [
    {
      "role": "user",
      "content": "What is Apple's P/E ratio?"
    },
    {
      "role": "assistant",
      "content": "Apple's P/E ratio is..."
    }
  ]
}
```

The streaming response is returned as newline-delimited JSON events:

```
{"type": "start", "request_id": "...", "question": "..."}
{"type": "node", "request_id": "...", "node": "decide_next_action", "output":
{...}}
{"type": "node", "request_id": "...", "node": "retrieve_knowledge_context", "out
put": {...}}
{"type": "node", "request_id": "...", "node": "generate_final_answer", "output":
 {...}}
{"type": "final", "request_id": "...", "response": {...}}
```

This coding structure supports a responsive and transparent user experience. The frontend does not only display the final answer; it also updates the trace as backend node events arrive. This allows users to observe whether the agent is retrieving knowledge, querying market data, analysing fundamentals, using web search, or generating the final answer.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261103

ESG Chatbot: An Intelligent Platform for HKEX ESG Analysis

Sergey VOLKOV [a], Frederick Nathan IRMAWAN [a], Lei LEI [a], Kunheng ZUO [a], Pai ZHANG [a], Duoyi WU [a], Yunqi AN [a] and Adela S.M. LAU[b,1]

[a] *Department of Statistics and Actuarial Science, School of Computing and Data Science, The University of Hong Kong, China*

[b] *Data Science Lab, Department of Statistics and Actuarial Science, School of Computing and Data Science, The University of Hong Kong, China*

ORCiD ID:
Sergey Volkov https://orcid.org/0000-0001-9379-6543
Frederick Nathan Irmawan https://orcid.org/0009-0001-4121-0405
Lei Lei https://orcid.org/0009-0008-9301-7587
Kunheng Zuo https://orcid.org/0009-0001-5892-6012
Pai Zhang https://orcid.org/0009-0001-1673-9083
Duoyi Wu https://orcid.org/0009-0002-9622-107X
Yunqi An https://orcid.org/0009-0007-7876-507X
Adela S.M. Lau https://orcid.org/0000-0001-5918-8309

Abstract. This project presents an end-to-end ESG intelligence platform for HKEX-listed companies, integrating natural language processing, deep learning, and retrieval-augmented generation to transform unstructured corporate disclosures into actionable insights. The platform adopts a two-tier analytical architecture: an offline batch processing layer employing lexicon-based topic classification, dictionary-based sentiment analysis, and TF-IDF keyword extraction for comprehensive signal extraction across 667 PDF reports, and an on-demand deep prediction layer built upon a fine-tuned BAAI bge-base-en-v1.5 embedding model that drives ESG score regression, multi-label topic classification, theme detection via semantic anchoring, and pillar-specific sentiment analysis. A conversational interface powered by GPT-4o-mini enables natural language querying with source-grounded citations. Results demonstrate that textual disclosure patterns contain predictive signals for ESG performance assessment, that semantic anchoring with fine-tuned embeddings effectively captures domain-specific ESG themes and sentiment across environmental, social, and governance dimensions, and that annual reports provide complementary disclosure evidence beyond standalone ESG reports.

Keywords. ESG, Natural Language Processing, Sentiment Analysis, ESG Score Prediction, Machine Learning, Hong Kong Stock Exchange

[1]Corresponding Author: Adela S.M. Lau, Data Science Lab, Department of Statistics and Actuarial Science, School of Computing and Data Science, Run Run Shaw Building, The University of Hong Kong, Pokfulam Road, Hong Kong; E-mail: adelalau@hku.hk.

1. Introduction

The Hong Kong Exchange (HKEX) mandates that all listed companies publish ESG reports [1][2], generating a vast and continuously expanding corpus of publicly available information. While this regulatory framework has significantly enhanced transparency, it has simultaneously created a critical bottleneck: investors and analysts now confront an overwhelming volume of unstructured textual data that resists systematic evaluation [3]. Although some researchers have developed the ESG instrument to evaluate company's ESG performance [4] or predict the ESG risks [5], manual review of hundreds of pages per company is resource-intensive, difficult to scale, and inherently subjective [6].

To address this gap, the present project develops an intelligent platform designed to automate the extraction, semantic processing, and quantitative analysis of ESG-related textual data from HKEX annual and ESG reports. The platform is constructed around a proprietary sentence embedding model fine-tuned with contrastive learning guided by external ESG ratings, enabling the transformation of unstructured text into a semantically meaningful vector space where companies with similar ESG profiles cluster together. Built upon this embedding foundation, the system performs keyword extraction, sentiment analysis, topic classification, and ESG score prediction using a suite of neural network-based modules. All analytical outputs are delivered through an interactive conversational chatbot interface, allowing users to query specific companies, compare disclosures across the market, and retrieve citation-backed answers in natural language.

The overarching objective of this project is to construct a domain-specific system that converts fragmented ESG narratives into structured, comparable, and actionable insights. Specifically, the project aims to:

- Develop an automated pipeline for parsing PDF-based reports from approximately 667 HK-listed companies;
- Fine-tune a sentence embedding model to capture ESG-specific semantic relationships;
- Implement NLP modules for keyword extraction, sentiment analysis, and topic classification;
- Construct and validate a deep MLP regressor for ESG score prediction;
- Integrate these components into an interactive RAG-based chatbot interface.

The scope of this study is confined to the textual analysis of English-language corporate disclosures and does not extend to real-time financial data integration or independent verification of reported ESG claims.

2. Literature review

2.1. Unstructured ESG Disclosures and the Need for Automated Extraction Pipelines

The rapid expansion of corporate sustainability reporting has resulted in an overwhelming volume of unstructured text. Mandatory disclosure regimes, most notably the Hong Kong Stock Exchange's (HKEX) enhanced climate-related reporting requirements effective from 1 January 2025, have forced thousands of listed companies to publish detailed Environmental, Social, and Governance (ESG) reports annually. Shen et al. [7] conducted a large-scale text analysis of over 210,000 annual reports from more

than 24,000 firms across 30 countries and found that while the volume of ESG disclosure increased considerably, the quality of information declined, becoming more templated and less firm-specific. These findings indicate that an analyst cannot simply read a few reports to assess a company's ESG posture; automated tools must process hundreds of documents at scale [7].

The first technical barrier lies in document ingestion. HKEX-listed companies release reports in diverse layouts that mix Chinese and English, embed complex tables, and frequently use scanned pages. Manually extracting text from such heterogeneous PDFs is error-prone. To address this, our project implements a dual-engine PDF extraction pipeline: PyMuPDF (fitz) [8] serves as the primary engine, with pdfplumber [9] as the fallback. This design choice is directly informed by the comprehensive benchmark conducted by Bast and Korzen [8], which demonstrated that no single PDF extraction tool performs optimally across all document types, and that combining tools can significantly improve robustness. When PyMuPDF fails to extract meaningful text---often because of complex table structures or non-standard encoding---pdfplumber's more lenient parser takes over, ensuring that the pipeline never loses an entire document due to formatting idiosyncrasies.

A second challenge is language mixing. Hong Kong corporate reports frequently contain both English and Chinese, but the downstream natural language processing (NLP) components we employ are pre-trained predominantly on English data. To maintain consistency, we implemented a regex-based Chinese character removal step that targets Unicode CJK blocks. This pragmatic filtering strategy follows the approach seen in cross-lingual ESG studies, where non-target languages are stripped before model ingestion. Although this sacrifices information contained in Chinese passages, it guarantees that the embedding and classification models operate on a homogeneous linguistic space, avoiding the distributional shift that would degrade performance.

After extraction, the plain text is cleaned by removing page headers, footers, and boilerplate phrases such as "HKEX" and "ANNUAL REPORT" that carry no ESG signal. The cleaned text is then segmented into overlapping chunks of 900 characters with a 120-character overlap. This configuration follows the dense passage retrieval guidelines established by Karpukhin et al. [11], who showed that a chunk size of approximately 100–200 tokens with meaningful overlap preserves semantic coherence while respecting the context window limits of modern embedding models. Each chunk is subsequently embedded using OpenAI's text-embedding-3-small model (1536 dimensions) [12] and stored in a vector database. This pre-computation forms the backbone of our offline bootstrap layer, enabling efficient, scalable retrieval across the entire corpus of 667 PDF documents without requiring on-the-fly LLM calls.

2.2. Large Language Models and Retrieval-Augmented Generation for ESG Structuring

The advent of large language models (LLMs) has opened new possibilities for extracting structured data from unstructured ESG narratives. Zou et al. [13] demonstrated the feasibility of an LLM-based framework, ESGReveal, that integrates Retrieval-Augmented Generation (RAG) to extract structured metadata from HKEX disclosures. By decomposing documents into manageable sections and applying targeted extraction prompts, ESGReveal achieved 83.7% accuracy with GPT-4. This work directly motivated our RAG-powered conversational chatbot, which allows users to ask natural language questions and receive answers grounded in the retrieved text chunks. We chose

to ground responses in source documents because, as George and Saji [14] demonstrated with their ESGBench benchmark, RAG-based models produce more specific, diverse, and factual language than parametric-only baselines, which is crucial for auditability.

The theoretical foundation of RAG was established by Lewis et al. [15], who combined a pre-trained parametric memory with a non-parametric memory (a dense vector index) to greatly improve performance on knowledge-intensive tasks. In our system, the parametric component is GPT-4o-mini [16], and the non-parametric memory is constructed from the offline layer's vector index. We set the generation temperature to zero to obtain deterministic, context-consistent answers, which is essential when users expect factual and comparable outputs across different queries and companies. This design echoes Senni et al. [17], who built the AskNature QA tool using RAG to assess corporate alignment with TNFD recommendations and stressed the importance of reproducibility in ESG evaluations.

Beyond answering queries, LLMs can also perform large-scale structuring of entire report corpora. Bronzini et al. [18] applied LLMs with in-context learning and RAG to extract structured insights from sustainability reports, discovering that ESG topics actually span over 500 distinct themes, far beyond conventional taxonomies. Their finding that ESG disclosure influences obtained ratings more than financial data reinforces our focus on text analytics. However, calling a commercial LLM for every report and every chunk would be economically prohibitive and slow. Therefore, our architecture separates concerns: the offline layer pre-computes embeddings and lightweight NLP features for all documents using inexpensive dictionary methods and a single embedding pass, while the on-demand deep layer applies the fine-tuned sentence transformer and MLP models only when a user requests a detailed analysis of a specific company. This two-tier approach mirrors the architectural pattern in modern RAG systems that decouple indexing from inference and ensure scalability across a dataset of 466 companies and 667 documents.

2.3. Sentence Transformers, Contrastive Fine-tuning, and Deep ESG Prediction

While the offline layer provides breadth, the deep layer requires semantic precision. To achieve this, we fine-tuned a BAAI bge-base-en-v1.5 sentence transformer [19]. This model was selected because of its state-of-the-art performance on information retrieval and semantic textual similarity benchmarks, and its 768-dimensional embedding space offers a good balance between expressive power and computational efficiency. The base checkpoint was then adapted to the ESG domain using contrastive learning, with HKQAA ratings serving as the supervisory signal. The contrastive learning paradigm, popularized in computer vision by Chen et al. [20] and extended to NLP by Gao et al. [21], works by pulling embeddings of similar samples together and pushing dissimilar ones apart. Wang et al. [22] were among the first to apply contrastive learning with BERT to ESG issue identification, showing that it enhances the model's ability to distinguish between semantically similar but distinct ESG topics. Our fine-tuning procedure follows this same principle: reports from companies with similar HKQAA ratings are treated as positive pairs in the contrastive loss, so the model learns ESG-relevant semantic features. The resulting fine-tuned retriever generates document embeddings that encode both the textual content and the latent ESG rating information, serving as the input to all deep-layer downstream modules.

The first downstream module is ESG score prediction. Many existing machine learning studies predict ESG ratings from financial and text data. Zhang et al. [23]

proposed the RMIDAS framework that integrates mixed-frequency data and found that Random Forest performed best for their heterogeneous dataset. Muhammad et al. [24] introduced a responsible AI framework using RuleFit to generate interpretable decision rules linking firm attributes to ESG scores. However, a systematic literature review by Seow [25] covering 57 high-quality articles from 2004–2024 concluded that common machine learning models systematically overweight quantifiable environmental metrics, underrepresenting the more qualitative social and governance dimensions. This bias arises because environmental data (e.g., carbon emissions) are often numeric and readily available, while social and governance information remains deeply embedded in text. To mitigate this bias, our deep regression model operates purely on the semantic content of the textual disclosures, without using any structured numeric inputs. We designed a 7-layer MLP regression head; the BatchNorm layers [26] and dropout [27] were applied throughout to regularize the training and prevent overfitting---a crucial consideration given our training set of 210 HK-listed companies with 5-fold cross-validation. The document vector is computed by encoding all chunks with the fine-tuned BAAI retriever, average-pooling them into a single 768-dimensional representation and then standardizing with a pre-fitted StandardScaler. The MLP outputs a continuous score between 0 and 100, which is subsequently mapped to the 12-level HKQAA rating scale, directly connecting qualitative text understanding with a quantitative, externally validated ESG rating.

2.4. Thematic Analysis: Prototypical Sentiment and Anchor-Based Topic Detection

Sentiment analysis in ESG reports requires a domain-specific approach because generic sentiment lexicons are often misleading. The widely used Loughran and McDonald (2011) financial sentiment dictionary, for instance, classifies "risk" as negative, but in an ESG context, "risk management" is a positive practice. In our offline layer, we implemented a lightweight, 20-word custom lexicon (10 positive and 10 negative terms) merely to provide a fast, baseline sentiment score for all documents. However, this coarse method cannot differentiate between E, S, and G dimensions, nor can it capture sentiment expressed through nuanced wording rather than single keywords.

To obtain dimensional sentiment signals with greater depth, the deep layer employs an unsupervised prototypical anchor mapping method. The core idea draws on Prototypical Networks [29], which classify instances by their proximity to class prototypes in an embedding space. Moodaley and Telukdarie [30] applied this concept to detect green claims in corporate text by mapping statements to prototypes of genuine versus misleading sustainability language. Adapting this for ESG sentiment, we manually crafted 10 positive and 10 negative prototype statements for each of the E, S, and G pillars---60 sentences in total, developed from domain knowledge and typical report language. For example, an E-positive prototype reads: "Achieved absolute reduction in greenhouse gas emissions across all scopes," while an E-negative prototype reads: "Significant increase in carbon intensity and total energy consumption." The fine-tuned BAAI retriever encodes both the document (via average-pooled chunk embeddings) and all prototypes. Then, for each pillar, the document vector's cosine similarity to the mean positive prototype vector and to the mean negative prototype vector is computed. The margin is compared against a threshold of ±0.03, yielding a positive, negative, or neutral label per ESG pillar. This prototype-based approach is entirely unsupervised with respect to sentiment labels---it requires no manually annotated sentiment data---and yet it yields dimensional, interpretable sentiment profiles

that align with the pillars of ESG reporting. Bronzini et al. [18] similarly used LLMs to extract structured insights from sustainability reports and noted that a wide range of topics beyond standard categories appear consistently; our anchor-based method provides the granularity to capture such diversity.

The deep layer also performs theme detection across 30 predefined ESG themes, such as Carbon Footprint, Energy Management, Waste Mitigation, and Sustainability Oversight. Each theme is represented by a descriptive anchor text, e.g., "Efforts to reduce greenhouse gas emissions across Scope 1, 2, and 3." Using the same fine-tuned BAAI retriever, we encode both every document chunk and the 30 theme anchors. A theme mention is recorded for a chunk whenever the cosine similarity between the chunk embedding and the anchor embedding exceeds 0.65. This threshold was empirically determined on a validation subset and provides a balance between recall (capturing genuine mentions) and precision (avoiding spurious matches). The result is a ranked list of the top 15 themes by mention frequency, offering a transparent and auditable thematic fingerprint of each company's report. In addition, a multi-label MLP classifier trained with binary cross-entropy predicts whether the document addresses seven broader topics: Community Involvement, Consumer Issues, Corporate Governance, Environment, Fair Operating Practices, Human Rights, and Labor Practices. This classifier is directly inspired by Schimanski et al. [31], who demonstrated that NLP models pre-trained on a large corpus of ESG text can effectively quantify E, S, and G communication and that a multi-label formulation captures overlapping themes better than single-label approaches. Together, the thematic detection modules provide structured visibility into the granular and high-level topics covered in corporate disclosures.

2.5. Research Gap and an Integrated End-to-End Platform

The preceding sections have surveyed a rich landscape of ESG analysis tools: BERT-based classifiers for GRI mapping [32][33], LLM-extractors like ESGReveal and CHATREPORT [13][34], RAG-based question answering systems such as ESGBench and AskNature [14][17], and lexical/topic models for disclosure measurement [31]. However, each of these systems addresses a single functional silo---extraction, classification, or question answering---and none provides a unified environment where an analyst can simultaneously explore top-level ESG scores, drill into thematic breakdowns, examine sentiment by pillar, and engage in a natural language dialogue about the evidence. He et al. [35] introduced ESGenius to benchmark LLMs on ESG knowledge and noted that real-world applications demand not just accurate model outputs but also usable interfaces that support human decision-making. This observation echoes the broader call for integrating data science methods into decision-support platforms.

Our project directly fills this gap by architecting a dual-layer system that marries scalable batch processing with on-demand deep semantic analysis, all accessible through an interactive web dashboard that includes a conversational chatbot. Architecture is a direct response to the challenges documented in literature. The offline bootstrap layer addresses the volume challenge [1],[7], by rapidly ingesting 667 PDFs, removing Chinese text, cleaning, chunking, and embedding all content with a cost-efficient embedding API. It also provides baseline NLP features---simple keyword-based topic classification, lexicon sentiment, TF-IDF keywords, and a rule-based ESG score---that populates the dashboard immediately, giving users a high-level map of the entire corpus. The on-demand deep layer, triggered when a user selects a specific company for detailed

analysis, employs the fine-tuned sentence transformer and the associated MLP regressor, prototype sentiment analyzer, and anchor-based theme detector to produce high-precision, multi-dimensional ESG profiles. This layer is specifically designed to overcome the biases and granularity limitations of the baseline methods and of many existing ML-based ESG scoring systems, which overfit environmental metrics [25]. The RAG chatbot, powered by the same embedding store and GPT-4o mini, ensures that every answer provided to the user is grounded in the retrieved report segments, making the system auditable and transparent, a requirement stressed by George and Saji [14] and Senni et al. [17].

In this way, the platform we have built is not a collection of isolated models but a coherent pipeline that translates the best practices from NLP, LLM, and contrastive learning research directly into a practical tool for HKEX ESG analysis. Each technical choice---from the dual-PDF extraction to the prototypical sentiment mapping---has a clearly traceable lineage in the academic literature, and together they form an integrated, scalable, and user-friendly solution that addresses the previously unmet need for an end-to-end ESG disclosure analysis system.

3. Data Science methods

3.1. Data Sources, Data Cleansing, and Pre-processing

Data Sources. The data corpus for this study comprises two distinct sources of information on HKEX-listed companies. The first source consists of structured JSON metadata records covering 466 companies, containing attributes including stock codes, industry classifications, official ESG ratings assigned by HKQAA, peer universe rankings, and documented corporate strengths and weaknesses. The second source comprises 667 unstructured PDF files, encompassing both standalone ESG reports and annual reports, covering a broader set of companies. The intersection of these two sources yields 210 companies for which both structured ratings and textual disclosures are available; this overlapping subset of 210 companies, for which both structured ratings and textual disclosures are available, provides the labeled foundation for training and fine-tuning the deep-layer models described below.

Data Cleansing and Extraction. Given that Hong Kong-listed companies frequently publish bilingual reports, a language filtering procedure was applied to isolate English-language content by removing Chinese character ranges, ensuring consistency for the English-optimized NLP components. PDF text extraction employed a dual-engine strategy: PyMuPDF as the primary extractor for high-fidelity retrieval, with pdfplumber as a fallback. Systematic noise removal followed, eliminating null bytes, page headers and footers, boilerplate phrases, and extraneous whitespace.

Pre-processing and Corpus Construction. Cleaned text was segmented into sentences using punctuation-delimited boundary detection, retaining sentences exceeding 25 characters. Documents were partitioned into overlapping chunks of 900 characters with a 120-character overlap, balancing semantic coherence with retrieval granularity. Two embedding strategies were employed: chunks destined for the RAG knowledge base were encoded into 1,536-dimensional vectors using OpenAI's text-embedding-3-small model and persisted in a vector-indexed database; in parallel, text for deep analytical models was encoded using a proprietary fine-tuned embedding model

based on the BAAI bge-base-en-v1.5 architecture, producing 768-dimensional vectors as the unified semantic representation for all downstream deep learning tasks.

3.2. Model Design

The analytical framework adopts a two-tier architecture: an offline batch processing layer for comprehensive, corpus-wide descriptive signal extraction, and an on-demand deep prediction layer for high-precision, company-specific inferential analysis. This section describes each layer in terms of its input and output variables, analytical methods, and role within the overall system.

3.2.1. Offline Layer: Descriptive Analytics and Signal Extraction

The offline layer processes all 667 PDF documents in a single batch execution, generating structured ESG signals that populate the platform's dashboards. The methods employed are lightweight, interpretable, and require no training data, making them suitable for full-corpus coverage.

Topic Classification. *Input:* individual sentences segmented from cleaned document text. *Method:* a lexicon-based classifier utilizing three domain-specific keyword sets of nine terms each, Environmental (e.g., emission, carbon, energy, climate), Social (e.g., employee, safety, diversity, inclusion), and Governance (e.g., board, audit, compliance, regulation). For each sentence, keyword frequencies are tallied per category using string containment matching; the highest-scoring category is assigned, with ties or absent matches defaulting to a Mixed label. *Output:* per-sentence topic label (E/S/G/Mixed), aggregated to per-document and per-company topic counts.

Threshold and label-assignment rationale. The offline topic classifier uses an intentionally transparent minimum-evidence rule: a sentence must contain at least one term from the Environmental, Social, or Governance lexicon to receive a pillar label; otherwise, it is assigned to Mixed. When more than one pillar receives the same highest keyword count, the sentence is also assigned to Mixed rather than forcing a potentially spurious E/S/G decision. This rule was chosen because the offline layer is designed for high-coverage descriptive analytics, not fine-grained supervised inference, and because conservative Mixed assignment reduces over-classification when sentences contain broad disclosure language.

Sentiment Analysis. *Input:* the subset of sentences classified as ESG-relevant (E, S, G, or Mixed). *Method:* a custom lexicon of 20 domain-relevant terms, 10 positive (e.g., improve, strong, efficient, effective) and 10 negatives (e.g., decline, penalty, breach, failure), scores each sentence as the difference between positive and negative word counts, normalized by total token count. Sentences with scores exceeding +0.015 are classified as positive, those below −0.015 as negative, and the remainder as neutral. *Output:* proportions of positive, neutral, and negative sentences per document, providing a unified sentiment profile across all ESG dimensions.

ESG Disclosure Density. *Input:* ESG-tagged sentence count and total sentence count per document. *Method:* a simple ratio computation. *Output:* a continuous density measure (0–1), serving as a standardized proxy for reporting comprehensiveness independent of document length.

Keyword Extraction. *Input:* cleaned document text. *Method:* term frequency-inverse document frequency (TF-IDF) weighting using unigram and bigram features

with English stop word removal, retaining the top 12 terms by mean TF-IDF score. *Output:* a ranked list of representative keywords per document.

Baseline ESG Rating Prediction. *Input:* aggregated E, S, G sentence counts and ESG disclosure density per company. *Method:* a rule-based weighted composition formula, with Governance weighted at 0.40, Environmental at 0.35, and Social at 0.25, reflecting the observed predominance of governance disclosures in the HKEX reporting landscape. The composition score is combined with density through a linear formula to produce a raw score mapped to an ordinal five-level rating scale. *Output:* a predicted rating category with an associated heuristic confidence score.

3.2.2. Deep Layer: Inferential Prediction Models

The deep layer is activated on demand when users request detailed predictions for a specific company. All models share a unified semantic backbone: a proprietary fine-tuned embedding model derived from BAAI bge-base-en-v1.5, producing 768-dimensional dense vectors. This model was fine-tuned using contrastive learning on the 210 companies with overlapping structured ratings and textual disclosures, employing the HKQAA ratings as guidance to pull embeddings of similarly rated companies closer in the semantic space while pushing those of differently rated companies apart. The shared embedding backbone ensures representational consistency across all deep analytical tasks.

ESG Score Prediction. *Input:* all text chunks belonging to a target company, encoded via the fine-tuned embedding model, mean-pooled to a single 768-dimensional document vector, and normalized through a pre-fitted standard scaler. *Method:* a deep multilayer perceptron regressor with progressively narrowing hidden layers, each employing batch normalization, ReLU activation, and dropout regularization with rates decreasing from 0.35 to 0.15 across layers. The model was trained on the 210-company overlapping subset using five-fold cross-validation, with the continuous score output mapped to a 12-tier HKQAA-aligned rating scale from AAA to BB. *Output:* a predicted ESG score (0–100), a mapped HKQAA rating category, and a confidence metric derived from the quantity of available text chunks and documents.

Fine-Grained Topic Classification. *Input:* the mean-pooled document embedding vector (768 dimensions) from the shared embedding backbone. *Method:* a multi-label MLP classifier with ReLU activation and dropout, trained to predict seven ESG topics: Community Involvement and Development, Consumer Issues, Corporate Governance, Environment, Fair Operating Practices, Human Rights, and Labor Practices. Sigmoid activation with a 0.5 threshold yields binary predictions. *Output:* per-topic probability scores and binary presence indicators.

Topic probability threshold rationale. The supervised multi-label classifier applies a 0.5 sigmoid threshold because each ESG topic is modeled as an independent binary decision; therefore, probabilities above 0.5 indicate that the model assigns greater evidence to presence than absence. This conventional threshold also keeps the decision rule consistent across all seven labels and was retained because the cross-validation results satisfied the predefined macro-F1 quality gate.

Theme Detection via Semantic Anchoring. *Input:* chunk-level embeddings and a curated set of 30 ESG sub-theme anchor descriptions (e.g., Carbon Footprint, Board Composition, Diversity and Inclusion, Waste Mitigation). *Method:* cosine similarity is computed between each chunk embedding and each theme anchor embedding;

associations exceeding a threshold of 0.65 are counted as theme mentions. *Output:* a ranked list of detected themes with mention frequencies, aggregated at the company level.

Theme similarity threshold rationale. The cosine-similarity cutoff of 0.65 was selected through an empirical validation sweep over candidate thresholds and manual inspection of retrieved theme matches. Lower values increased recall but admitted generic sustainability language as false positives, while higher values improved precision at the cost of missing legitimate paraphrases of the same ESG theme. The 0.65 cutoff was therefore adopted as a practical balance between precision, recall, and interpretability for the 30 anchor-based themes.

E/S/G-Dimensional Sentiment Analysis. *Input:* the mean-pooled document embedding vector and 60 prototypical sentiment anchor statements, 10 positive and 10 negative per E, S, and G pillar, encoded through the same embedding backbone and averaged per pillar-polarity combination. *Method:* cosine similarity is computed between the document vector and each prototype mean vector; a sentiment margin (positive similarity minus negative similarity) is calculated per pillar, with a threshold of ±0.03 determining polarity. *Output:* independent sentiment labels (positive, neutral, or negative) for each of the E, S, and G dimensions

3.2.3. Retrieval-Augmented Generation Interface

The conversational interface employs a hybrid retrieval mechanism combining semantic similarity (cosine similarity between query and chunk embeddings, weight 0.7) with lexical overlap (token-level Jaccard similarity, weight 0.3). Retrieved chunks with source metadata are assembled as context for GPT-4o-mini (temperature=0), which generates responses grounded in the provided evidence with inline citations. Five quantitative quality gates, topic macro-F1 $\geq$0.78, prediction MAE $\leq$1.0, citation coverage $\geq$0.95, Groundedness $\geq$0.85, and P95 latency $\leq$8 seconds, serve as automated evaluation thresholds.

3.2.4. Data Visualization

A multi-view dashboard presents aggregate market-level statistics, individual company profiles with ESG signal breakdowns and predictions, and a comparative analysis view enabling side-by-side company evaluation along standardized ESG dimensions with density-based ranking and recommendation summaries.

4. Result summary and Interpretation

4.1. Corpus Overview and Descriptive Statistics

The analytical platform ingests two complementary data sources: structured ESG ratings in JSON format covering 466 HKEX-listed companies, and 667 unstructured PDF documents comprising both standalone ESG reports and annual reports. The intersection of these two sources yields 210 companies for which both ground-truth ratings and textual disclosures are available, forming the supervised training set for the fine-tuned embedding model. From the full document corpus, the offline processing pipeline successfully extracted valid ESG signals for 660 companies, with the remaining documents excluded due to insufficient extractable textual content.

To characterize the corpus at a high level, the aggregate distribution of ESG topics across all extracted sentences was examined. Figure 1 presents the total sentence counts for each of the four topic categories assigned by the lexicon-based classifier. Governance-related sentences formed the largest category, reflecting the mandatory disclosure requirements under HKEX listing rules that compel detailed reporting on board structures, audit committees, and risk management frameworks. Environmental and Social sentences appeared at lower but broadly comparable frequencies, while the Mixed category, assigned when a sentence contained no dominant keyword or matched multiple ESG dimensions equally, registered the smallest share. This topic hierarchy is consistent with the compliance-driven nature of corporate disclosure in Hong Kong, where governance reporting obligations precede and exceed those for environmental and social matters.

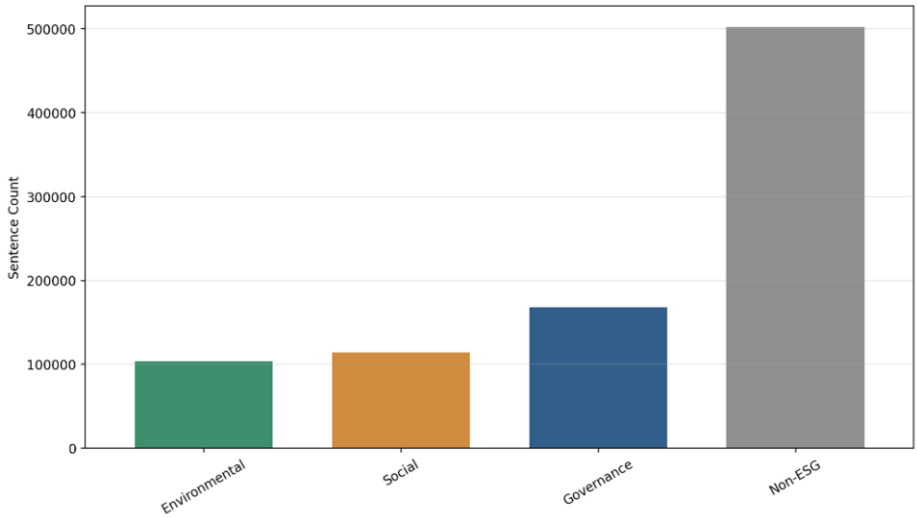


Figure 1. ESG Topic Distribution Across Corpus

Beyond raw topic counts, the platform computes an ESG disclosure density metric, defined as the proportion of ESG-tagged sentences (E, S, and G categories, excluding Mixed to avoid inflating the measure with ambiguous content) relative to total sentences, as a standardized indicator of reporting comprehensiveness. Table 1 summarizes the density distribution across the 660 companies with valid measurements. The mean density of 0.468 (SD=0.111) and median of 0.505 indicate that, on average, approximately half of the textual content in these corporate documents addresses ESG-related topics. The relatively narrow interquartile range (Q1=0.354, Q3=0.554) suggests a modest degree of variation in disclosure intensity across the HKEX-listed universe.

Table 1. ESG Disclosure Density Summary Statistics (N=660)

Statistic	Value
Mean	0.4681
Median	0.5052
Standard Deviation	0.1109

Minimum	0.0570
Maximum	0.6760

A more detailed inspection of the density distribution, presented as a histogram in Figure 2, reveals a distinctive bimodal structure. The primary mode, encompassing 304 companies (46.1%), lies in the 0.50–0.60 range, while a secondary mode of 174 companies (26.4%) occupies the 0.30–0.40 range. This bimodality suggests the presence of two distinct reporting populations: companies in the higher-density mode likely produce ESG-dedicated or sustainability-focused documents where ESG content is inherently concentrated, whereas those in the lower-density mode appear to embed ESG disclosures within broader annual reports, where such content is interspersed with financial results and operational commentary. Supporting this interpretation, companies at the extreme lower tail, below 0.20 density, account for fewer than 1% of the sample (3 companies), suggesting that the regulatory impetus provided by HKEX's mandatory ESG reporting requirements has driven near-universal adoption of substantial ESG disclosure practices.

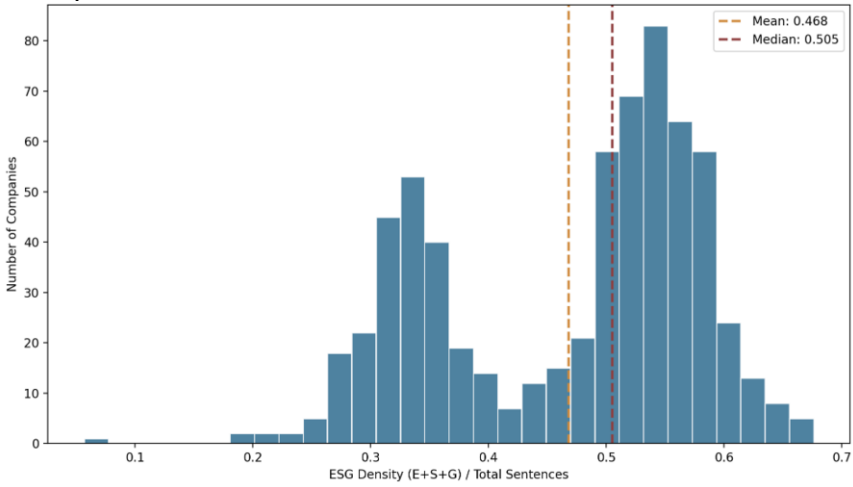


Figure 2. Distribution of ESG Disclosure Density Across Companies

A further noteworthy feature of the density distribution is the observed ceiling effect: no company exceeded a density of 0.676. This upper bound indicates that even in the most sustainability-focused documents, a meaningful proportion of textual content remains dedicated to non-ESG matters such as corporate strategy, financial highlights, and administrative disclosures. Combined with the narrow overall spread, this ceiling reinforces the characterization of ESG disclosure density as a relatively standardized practice across the HKEX-listed universe, bounded at both the lower and upper extremes by regulatory expectations and the inherent multifunctionality of corporate reporting documents, respectively.

4.2. Retriever Fine-Tuning and ESG Score Prediction Results

The deep-layer ESG score prediction comprises two interconnected stages: fine-tuning the embedding model that serves as the unified semantic backbone and training the ESG score predictor that maps document representations to rating scores.

Retriever Fine-Tuning. The BAAI bge-base-en-v1.5 embedding model was fine-tuned using the 210 companies for which both HKQAA ratings and sufficient textual disclosures were available. Training pairs were constructed by sampling text chunks from companies with similar ratings (within one notch; positive pairs) and from companies with substantially different ratings (four or more notches apart; negative pairs). A cosine similarity loss function was applied, pulling embeddings of similarly rated companies closer in the semantic space while pushing those of differently rated companies apart. This contrastive learning strategy encourages the embedding space to organize along an ESG performance dimension, such that companies with comparable sustainability profiles yield similar vector representations regardless of the specific vocabulary used in their disclosures.

ESG Score Predictor. The fine-tuned retriever encodes all text chunks belonging to a target company into 768-dimensional vectors, which are mean-pooled to produce a single document-level embedding. A deep MLP regressor with batch normalization, ReLU activation, and graduated dropout then maps this embedding to a continuous ESG score (0–100), subsequently converted to a 12-tier HKQAA-aligned rating.

Fine-Tuning Impact. To isolate the contribution of the fine-tuned retriever, the identical regression architecture was trained and evaluated under two conditions: using the original BAAI bge-base-en-v1.5 embeddings, and using the fine-tuned embeddings. Both conditions employed the same 210-company subset, five-fold cross-validation, and identical training hyperparameters. Table 2 presents the comparative results.

Table 2. ESG Score Prediction Performance: Before vs. After Retriever Fine-Tuning

Metric	Original BAAI	Fine-Tuned BAAI	Improvement
MAE	10.32	**6.92**	↓ 33.0%
R^2	0.56	**0.80**	↑ 42.9%
Exact Match	23.1%	**36.2%**	↑ 13.1%
Within ±1 Tier	69.7%	**86.7%**	↑ 17.0%

The fine-tuned retriever yielded substantial improvements across all metrics. Mean absolute error decreased from 10.32 to 6.92, representing a 33% reduction in prediction error. The R^2 increased from 0.56 to 0.80, indicating that the fine-tuned embeddings capture substantially more variance in ESG performance. In terms of rating classification, exact match accuracy improved from 23.1% to 36.2%, and the within-one-tier tolerance, a practically meaningful measure allowing adjacent rating categories, rose from 69.7% to 86.7%. These results demonstrate that domain-adaptive fine-tuning of the embedding model, guided by the very ratings the system aims to predict, provides a materially

stronger semantic foundation for downstream ESG assessment than off-the-shelf general-purpose embeddings.

The trained predictor is deployed through the platform's prediction interface, where users can trigger on-demand ESG score computation for any company. As illustrated in Figure 3, the dashboard displays the predicted rating, confidence score, model version, and the number of chunks and documents processed, providing full traceability from raw disclosures to final assessment.

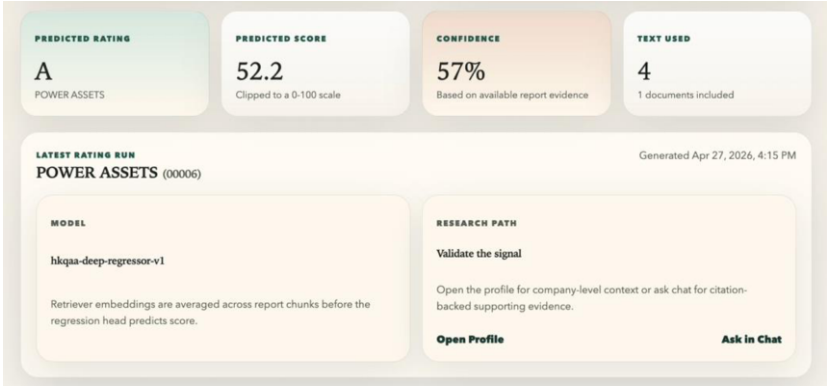


Figure 3. ESG Score Prediction Dashboard for the company Power Assets

4.3. Theme Detection Results

The theme detection module identifies the presence of 30 predefined ESG sub-themes within company disclosures using semantic anchoring. Each sub-theme, ranging from Carbon Footprint and Waste Mitigation to Board Composition and Diversity and Inclusion, is represented by a natural language anchor description. Both the anchor descriptions and each text chunk from a company's documents are encoded into 768-dimensional vectors using the fine-tuned retriever. A theme is considered detected in a given chunk when the cosine similarity between the chunk embedding and the anchor embedding exceeds 0.65. Mentions are then aggregated across all chunks and ranked by frequency.

Figure 4 presents the dashboard output of this analysis for the company CK Hutchison Holdings (Stock code: 00001), drawn from 585 text chunks across its available ESG reports and annual reports. The horizontal bars display detected themes ranked by the number of chunks in which each theme was identified.



Figure 4. Theme Detection Dashboard for the company CK Hutchison Holdings

The results reveal broad thematic coverage spanning all three ESG pillars. Governance-adjacent themes, Sustainability Oversight (377 chunks), Executive Pay (373), and Political Engagement (360), featured prominently, consistent with the governance-heavy composition of HKEX-mandated disclosures observed in the aggregate corpus distribution. Environmental themes registered comparable presence, with Waste Mitigation (378), Carbon Footprint (376), and Energy Management (342) all exceeding 300 mentions, reflecting the materiality of climate-related disclosures. Social themes including Diversity and Inclusion (373) and Health and Safety (325) also ranked among the top entries, indicating balanced coverage across E, S, and G dimensions.

Beyond the top-ranked themes, the distribution extended to a broader set of topics including Privacy and Data (229), Community Impact (194), Business Ethics (194), Climate Strategy (158), and Risk Management (152). The presence of these secondary themes at lower but non-trivial frequencies suggests that the semantic anchoring approach captures fine-grained thematic nuances rather than merely identifying dominant disclosure categories. The threshold mechanism, requiring cosine similarity above 0.65, ensures that detected mentions reflect genuine thematic alignment rather than incidental keyword overlap.

4.4. Fine-Grained Topic Classification Results

The topic classification module performs supervised multi-label prediction across seven ESG topics: Community Involvement and Development, Consumer Issues, Corporate Governance, Environment, Fair Operating Practices, Human Rights, and Labor Practices. The classifier employs an MLP head (768→512→128→7) trained on labelled data, taking as input the same mean-pooled document embedding produced by

the fine-tuned retriever. A sigmoid activation function outputs independent probability scores for each topic, with values exceeding 0.5 indicating a positive prediction. Unlike the coarse E/S/G/Mixed classifier in the offline layer, this model distinguishes semantically distinct sub-dimensions within each ESG pillar, for instance, separating Labor Practices and Human Rights within the Social domain.

Figure 5 presents the dashboard output for the company Power Assets (Stock code: 00006).

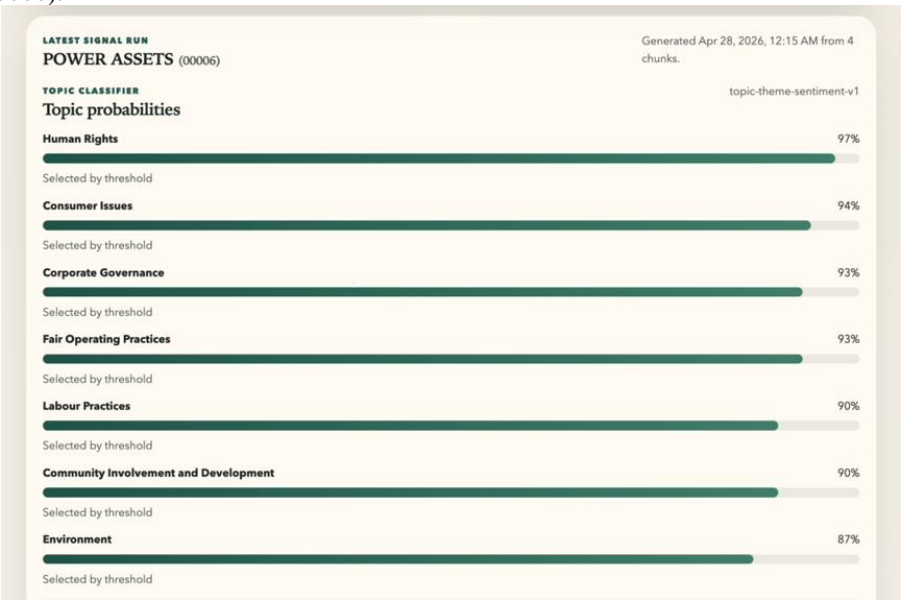


Figure 5. Topic Classification Dashboard for the company Power Assets

All seven topics exceeded the 0.5 threshold, indicating that the disclosures of Power Assets substantively address the full spectrum of ESG topics. Human Rights, Consumer Issues, and Corporate Governance received the highest probabilities (97%, 94%, and 93% respectively), while Environment registered the lowest at 87%. The consistently high probabilities across all categories reflect the classifier's confidence that each topic is addressed in the company's reporting, though the relative ranking provides a coarse indication of disclosure emphasis. This multi-label output complements the theme detection results: where theme detection identifies specific sub-themes through unsupervised semantic matching, topic classification provides a supervised assessment of whether broader ESG categories are substantively covered.

4.5. E/S/G-Dimensional Sentiment Analysis Results

The deep-layer sentiment analysis module produces independent sentiment assessments for each ESG pillar using prototypical anchor mapping. For each of the Environmental, Social, and Governance dimensions, 10 positive and 10 negative prototype statements were constructed to embody ideal sentiment polarities within that domain, for instance, "Achieved absolute reduction in greenhouse gas emissions across all scopes" (E-positive) versus "Significant increase in carbon intensity and total energy consumption" (E-negative). All 60 prototypes are encoded through the fine-tuned retriever and averaged per pillar-polarity combination to form six reference vectors. A

company's mean-pooled document embedding is compared against these reference vectors via cosine similarity; the sentiment margin, positive similarity minus negative similarity, determines polarity, with a threshold of ±0.03 separating positive, neutral, and negative labels. This design enables the detection of sentiment divergence across ESG dimensions that a unified sentiment score would obscure.

Figure 6 presents the dashboard output for the company CK Hutchison Holdings.



Figure 6. E/S/G Sentiment Analysis Dashboard for the company CK Hutchison Holdings

All three pillars returned positive sentiment labels. The Environmental dimension exhibited the strongest positive margin (0.09), with the document embedding showing higher similarity to environmental positive prototypes (0.80) than to negative ones (0.71). The Social dimension followed with a margin of 0.08 (positive similarity 0.76, negative 0.68), and Governance registered a margin of 0.06 (positive 0.75, negative 0.68). The uniformly positive but variably strong margins suggest that while CK Hutchison Holdings' disclosures convey a positive tone across all ESG dimensions, the degree of positivity varies, with environmental disclosures expressing the most favorable sentiment and governance disclosures the least. This pattern is consistent with the tendency for environmental reporting to emphasize achievements and targets, while governance reporting, constrained by compliance-oriented language, maintains a comparatively measured tone.

4.6. RAG Conversational Interface Evaluation

The retrieval-augmented generation system enables users to query the ESG knowledge base through natural language. User questions are encoded via OpenAI's text-embedding-3-small model and matched against the pre-indexed chunk vector database using a hybrid scoring mechanism that combines semantic similarity (cosine similarity, weight 0.7) with lexical overlap (token-level Jaccard similarity, weight 0.3). Retrieved chunks with source metadata are assembled as context for GPT-4o-mini, which generates responses grounded in the provided evidence with inline citations.

The system was evaluated against five quantitative performance gates using a set of representative ESG queries. Table 3 presents the results.

Table 3. RAG System Evaluation Against Quality Gates

Metric	Threshold	Achieved	Status
Topic Macro-F1	≥ 0.78	0.80	✓ Pass
Citation Coverage	≥ 0.95	1.00	✓ Pass

Metric	Threshold	Achieved	Status
Groundedness	≥ 0.85	1.00	✓ Pass
Prediction MAE	≤ 1.0	0.90	✓ Pass
P95 Latency	≤ 8s	11.78s	✗ Fail

Citation coverage and Groundedness both achieved perfect scores of 1.00, indicating that all evaluated responses included source references and that generated answers were fully supported by retrieved evidence. The topic classification macro-F1 of 0.80 met its threshold of 0.78, and the ESG prediction MAE of 0.90 satisfied the ≤1.0 requirement. The P95 latency of 11.78 seconds exceeded the 8-second target, representing the sole metric below threshold. This latency encompasses the full query pipeline, query embedding generation, vector similarity search, context assembly, and LLM inference across a corpus of 667 documents, and identifies response speed as the primary area for future optimization.

4.7. Conversational Interface and Dashboard Visualisation

The platform integrates a FastAPI backend serving REST endpoints with a Next.js frontend delivering the user interface. All ESG signals, predictions, and retrieval results computed by the offline and deep-layer pipelines are exposed through these endpoints and rendered across multiple interactive views. Figure 7 presents the conversational interface, where users submit natural language queries and receive grounded responses with inline citations referencing source documents, stock codes, and report types. Each citation is traceable to its originating chunk, enabling users to verify claims against original disclosures.

In addition to the conversational interface, the platform provides three further views that have been illustrated through figures in preceding sections: the theme detection dashboard (Figure 4), the topic classification panel (Figure 5), and the E/S/G-dimensional sentiment analysis display (Figure 6), all accessible from the company profile page. Complementing these, a comparative analysis view enables side-by-side evaluation of multiple companies along standardized ESG dimensions, disclosure density, dominant topic, sentiment balance, and rating, with density-based ranking and automated recommendation summaries identifying the strongest ESG disclosure profile among the selected set. An overview dashboard further aggregates market-level statistics including company counts by industry, average ESG density, and aggregate topic and sentiment distributions. Together, these interfaces transform the underlying structured signals and deep predictions into interpretable decision-support tools for stakeholders ranging from investors to regulators.

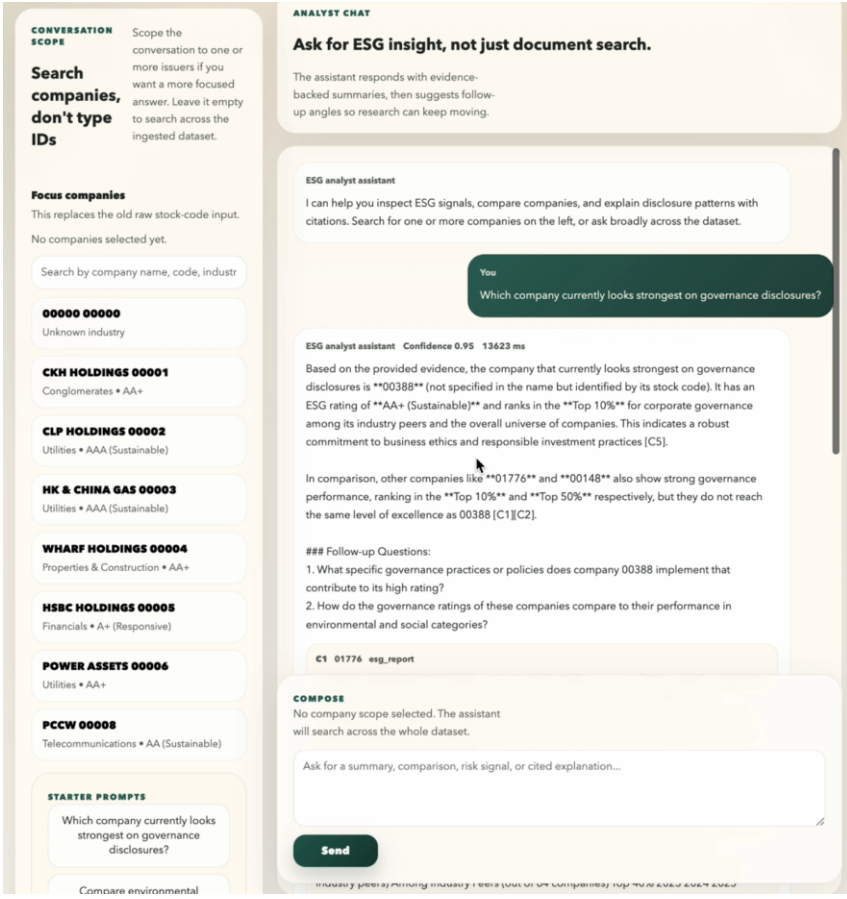


Figure 7. RAG Conversational Interface

5. Discussions

The results presented in the preceding sections demonstrate that the platform achieves its core objective of transforming unstructured ESG disclosures into structured, actionable intelligence. This section interprets the key findings and explains how the employed methods fulfil the project goals.

Effectiveness of the Proprietary Embedding Model. The fine-tuned BAAI embedding model served as the unified semantic backbone for four distinct analytical tasks, ESG score prediction, theme detection, topic classification, and dimensional sentiment analysis, demonstrating the versatility of domain-adapted embeddings for multi-task ESG intelligence. The retriever fine-tuning, guided by HKQAA ratings through contrastive learning, yielded substantial improvements: MAE declined from 10.32 to 6.92 and within-one-tier accuracy rose from 69.7% to 86.7%. This finding indicates that general-purpose embeddings, despite broad linguistic competence, lack sensitivity to the domain-specific semantic distinctions that differentiate ESG performance levels. The fine-tuning process effectively reorganized the embedding

space along an ESG performance dimension, enabling the same downstream architecture to extract substantially more predictive signal from identical input texts.

Validity of the Semantic Anchoring Approach. Both the 30-theme detection and the E/S/G-dimensional sentiment analysis employed semantic anchoring, comparing document embeddings against curated prototype vectors in the shared embedding space, and produced coherent, interpretable results on real-world corporate disclosures. The theme detection module identified granular sub-themes with a distribution pattern consistent with known characteristics of HKEX reporting, while the dimensional sentiment analysis revealed meaningful divergence across ESG pillars that a unified sentiment measure would obscure. These findings validate the premise that well-designed anchor descriptions, when encoded in a domain-adapted embedding space, can serve as effective zero-shot classifiers for nuanced ESG concepts without requiring extensive labelled training data for each new theme or sentiment dimension.

Methods Achieving Project Goals. The project aimed to address information overload and lack of actionable intelligence by automating the extraction, processing, and analysis of ESG data. The two-tier architecture directly serves this goal: the offline layer provides immediate, corpus-wide coverage for dashboard-based comparison, while the deep layer delivers high-precision, on-demand analysis for granular due diligence. The RAG conversational interface further lowers the barrier to insight extraction by enabling natural language queries with traceable, cited responses. Together, these components transform what would otherwise require manual review of hundreds of pages of disclosures into an interactive, query-driven analytical workflow.

6. Limitations

The proposed platform should be interpreted as a decision-support system rather than a definitive ESG rating authority. Several limitations follow directly from the methodological choices made to keep the system scalable, interpretable, and applicable across a large HKEX disclosure corpus.

First, the analysis is restricted to English-language text. Chinese passages were removed to maintain compatibility with the English-optimized embedding, classification, and generation components. This improves linguistic consistency but may discard relevant information from bilingual reports, especially for companies that provide more detailed sustainability explanations in Chinese.

Second, the pipeline relies on extracted PDF text and self-reported corporate disclosures. Although the dual-engine extraction strategy improves robustness, scanned pages, complex tables, and unusual layouts may still introduce extraction noise. In addition, the system analyzes what companies disclose; it does not independently verify whether reported ESG claims are complete, externally audited, or free from greenwashing.

Third, the supervised deep-layer models are trained on the 210 companies for which both textual disclosures and HKQAA ratings are available. This labeled subset provides a useful foundation, but it may not fully represent all industries, reporting styles, and rating distributions in the broader HKEX universe. The resulting ESG score predictions should therefore be used comparatively and with caution rather than as standalone investment advice.

Fourth, several analytical components use manually designed lexicons, prototype statements, theme anchors, and validation-tuned thresholds. These choices improve

transparency and interpretability, but they may require recalibration when the platform is applied to other markets, languages, reporting years, or ESG taxonomies. Future work should test threshold robustness across additional annotated datasets and expand the theme inventory beyond the 30 anchors used in this study.

Finally, the RAG interface depends on commercial embedding and language-model services and was evaluated using representative ESG queries rather than a full production workload. Model updates, API costs, latency constraints, and prompt sensitivity may affect reproducibility over time. Future versions should include broader human evaluation, stronger multilingual support, and more systematic stress testing under production-scale usage.

7. Conclusions

This project developed an integrated platform that automates the extraction, processing, and analysis of ESG disclosures from HKEX-listed companies, transforming unstructured reports into structured signals, predictive ratings, and an interactive conversational interface. Several gaps between this work and existing approaches merit acknowledgement.

First, Bronzini et al. [18] demonstrated that ESG topics span over 500 distinct themes, far exceeding the 30 sub-themes and seven categories currently covered by our platform. For future work, the project could expand its anchor library to include emerging themes such as biodiversity, supply chain ethics, and just transition, enabling more comprehensive thematic coverage.

Second, the project's RAG system achieved a P95 latency of 11.78 seconds, exceeding the 8-second target established in our evaluation framework. Future iterations of the platform could incorporate embedding caching strategies and vector index optimization through approximate nearest neighbor search to improve response times under increasing document volumes.

Finally, Shen et al. [7] observed that as ESG disclosure volume increases, disclosure quality often declines. Our platform currently treats all disclosed content equally; future work could integrate a disclosure quality assessment module that distinguishes substantive, firm-specific reporting from generic boilerplate, strengthening the platform's evaluation reliability.

References

[1] HKEX. (2024, April 19). Consultation conclusions on enhancement of climate-related disclosures under the Environmental, Social and Governance framework. Hong Kong Exchanges and Clearing Limited. https://www.hkex.com.hk/News/Regulatory-Announcements/2024/240419news

[2] HKEX. (2025). Appendix C2: Environmental, Social and Governance Reporting Code. Hong Kong Exchanges and Clearing Limited. https://en-rules.hkex.com.hk/rulebook/appendix-c2-environmental-social-and-governance-reporting-code-0

[3] Krueger, P., Sautner, Z., & Starks, L. T. (2020). The importance of climate risks for institutional investors. The Review of Financial Studies, 33(3), 1067–1111. https://doi.org/10.1093/rfs/hhz137

[4] Lau ASM, Liu H, Hung V, Li C, Lam M. Development of a Novel Integrated Ontology-Based ESG Assessment Tool with AI Assistance for SMEs. Frontier in Artificial Intelligence and Applications. 2025;412:369-391. http://dx.doi.org/10.3233/FAIA250736

[5] Dong YX, Lau ASM. A Review of Statistical and AI Methods for Predicting ESG Risks for Default. Frontier in Artificial Intelligence and Applications. 2025;412:317-321. https://doi.org/10.3233/FAIA250731
[6] Berg, F., Koelbel, J. F., & Rigobon, R. (2022). Aggregate confusion: The divergence of ESG ratings. Review of Finance, 26(6), 1315–1344. https://doi.org/10.1093/rof/rfac033
[7] Shen, R., Lin, Y., Wang, J., & Yu, Y. J. (2025). Global evolution of environmental and social disclosure in annual reports. Journal of Accounting Research, 63(2), 345–398. https://doi.org/10.1111/1475-679X.12567
[8] McKie, J. (2023). PyMuPDF: Python binding for MuPDF (Version 1.23.x) [Computer software]. https://pymupdf.readthedocs.io/
[9] Singer, J. (2023). pdfplumber: Plumb a PDF for detailed information about each text character, rectangle, and line (Version 0.10.x) [Computer software]. https://github.com/jsvine/pdfplumber
[10] Bast, H., & Korzen, C. (2017). A benchmark and evaluation for text extraction from PDF. In J. S. Downie (Ed.), Proceedings of the 17th ACM/IEEE Joint Conference on Digital Libraries (JCDL) (pp. 99–108). IEEE.
[11] Karpukhin, V., Oğuz, B., Min, S., Lewis, P., Wu, L., Edunov, S., Chen, D., & Yih, W. (2020). Dense passage retrieval for open-domain question answering. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (pp. 6769–6781). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.emnlp-main.550
[12] OpenAI. (2024a, January 25). New embedding models and API updates. https://openai.com/index/new-embedding-models-and-api-updates/
[13] Zou, Y., Shi, M., Chen, Z., Deng, Z., Lei, Z., Zeng, Z., Yang, S., Tong, H., Xiao, L., & Zhou, W. (2025). ESGReveal: An LLM-based approach for extracting structured data from ESG reports. Journal of Cleaner Production, 489, Article 144572. https://doi.org/10.1016/j.jclepro.2024.144572
[14] George, S. V., & Saji, N. (2025). ESGBench: A benchmark for explainable ESG question answering in corporate sustainability reports. arXiv. https://arxiv.org/abs/2511.16438
[15] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. In Advances in Neural Information Processing Systems 33 (NeurIPS 2020) (pp. 9459–9474). https://proceedings.neurips.cc/paper/2020/file/6b493230205f780e1bc26945df7481e5-Paper.pdf
[16] OpenAI. (2024b, July 18). GPT-4o mini: Advancing cost-efficient intelligence. https://openai.com/index/gpt-4o-mini-advancing-cost-efficient-intelligence/
[17] Senni, C. C., Vaghefi, A. S., Manekar, T., Schimanski, T., & Leippold, M. (2025). Using AI to assess the decision-usefulness of corporates' nature-related disclosures. Environmental Research Communications, 7(11), Article 115006. https://doi.org/10.1088/2515-7620/ae1357
[18] Bronzini, M., Nicolini, C., Lepri, B., Passerini, A., & Staiano, J. (2024). Glitter or gold? Deriving structured insights from sustainability reports via large language models. EPJ Data Science, 13(1), Article 41. https://doi.org/10.1140/epjds/s13688-024-00481-2
[19] Xiao, S., Liu, Z., Zhang, P., & Muennighoff, N. (2024). C-pack: Packed resources for general Chinese embeddings. In E. Yom-Tov (Ed.), Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 641–649). Association for Computing Machinery. https://doi.org/10.1145/3626772.3657875
[20] Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A simple framework for contrastive learning of visual representations. In Proceedings of the 37th International Conference on Machine Learning (ICML) (pp. 1597–1607). PMLR.
[21] Gao, T., Yao, X., & Chen, D. (2021). SimCSE: Simple contrastive learning of sentence embeddings. In M.-F. Moens, X. Huang, L. Specia, & S. W. Yih (Eds.), Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (pp. 6894–6910). Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.emnlp-main.552
[22] Wang, W., Wei, W., Song, Q., & Wang, Y. (2023). Leveraging contrastive learning with BERT for ESG issue identification. In Proceedings of the Fifth Workshop on Financial Technology and Natural Language Processing and the Second Multimodal AI for Financial Forecasting (pp. 116–120). Association for Computational Linguistics. https://aclanthology.org/2023.finnlp-1.12.pdf
[23] Zhang, Y., Ding, X., Wang, L., & Shi, Y. (2026). Does mixed-frequency data efficiently predict future ESG ratings? A RMIDAS-based machine learning approach. Business Strategy and the Environment, 35(3), 1–25. https://doi.org/10.1002/bse.70884
[24] Muhammad, T., Hafeez, R., Khidmat, W. B., Afzal, M., Abid, F., Elkiran, H., Gurhanli, A., & Rasheed, J. (2026). A hypothesis-driven responsible AI framework for interpretable ESG forecasting with RuleFit. Scientific Reports, 16, Article 40562. https://doi.org/10.1038/s41598-026-47870-3

[25] Seow, R. Y. C. (2025). Transforming ESG analytics with machine learning: A systematic literature review using TCCM framework. Corporate Social Responsibility and Environmental Management, 32(6), 7358–7389. https://doi.org/10.1002/csr.70089

[26] Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the 32nd International Conference on Machine Learning (pp. 448–456). PMLR. https://doi.org/10.48550/arXiv.1502.03167

[27] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 15(1), 1929–1958. https://www.jmlr.org/papers/v15/srivastava14a.html

[28] Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. The Journal of Finance, 66(1), 35–65. https://doi.org/10.1111/j.1540-6261.2010.01625.x

[29] Snell, J., Swersky, K., & Zemel, R. (2017). Prototypical networks for few-shot learning. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), Advances in Neural Information Processing Systems 30 (NeurIPS 2017) (pp. 4077–4087).

[30] Moodaley, W., & Telukdarie, A. (2023). A conceptual framework for subdomain specific pre-training of large language models for green claim detection. European Journal of Sustainable Development, 12(4), 319. https://doi.org/10.14207/ejsd.2023.v12n4p319

[31] Schimanski, T., Reding, A., Reding, N., Bingler, J., Kraus, M., & Leippold, M. (2024). Bridging the gap in ESG measurement: Using NLP to quantify environmental, social, and governance communication. Finance Research Letters, 61, Article 104979. https://doi.org/10.1016/j.frl.2024.104979

[32] Hillebrand, L., Pielka, M., Leonhard, D., Deußer, T., Dilmaghani Khameneh, T., Kliem, B., Loitz, R., Morad, M., Temath, C., Bell, T., Stenzel, R., & Sifa, R. (2023). sustain.AI: A recommender system to analyze sustainability reports. In Proceedings of the Nineteenth International Conference on Artificial Intelligence and Law (ICAIL 2023) (pp. 412–416). Association for Computing Machinery. https://doi.org/10.1145/3594536.3595131

[33] Ruberg, P. (2023). BERT goes sustainable: A multilingual approach to GRI classification in ESG reports [Master's thesis, Pontifical Catholic University of Rio de Janeiro].

[34] Ni, J., Bingler, J., Colesanti Senni, C., Kraus, M., & Leippold, M. (2023). CHATREPORT: Democratizing sustainability disclosure analysis through LLM-based conversational tools. In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: System Demonstrations (pp. 138–145). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.emnlp-demo.3

[35] He, C., Zhou, X., Wu, Y., Yu, X., Zhang, Y., Zhang, L., Wang, D., Lyu, S., Xu, H., Wang, X., Liu, W., & Miao, C. (2025). ESGenius: Benchmarking LLMs on environmental, social, and governance (ESG) and sustainability knowledge. In Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing (pp. 14612–14653). Association for Computational Linguistics. https://doi.org/10.18653/v1/2025.emnlp-main.739

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)

© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261104

A Retrieval-based Multi-Source Medical Chatbot: Integrating Association Rule Mining and Bayesian Inference as a Health Information Support Tool

Peilin LUO[a], Anmaria ALICE [a], Zongfu CHEN [a], Jiayin GUO [c], Xinchu ZHOU [a], Yutong GU [a] and Adela S. M. LAU [b,1]
[a] *School of Computing and Data Science, The University of Hong Kong, China*
[b] *Data Science Lab, Department of Statistics and Actuarial Science, School of Computing and Data Science, The University of Hong Kong, China*
[c] *Department of Chemistry, The University of Hong Kong, China*
ORCiD ID:
Peilin LUO https://orcid.org/0009-0005-2251-9959
Anmaria ALICE https://orcid.org/0009-0003-4794-1625
Zongfu CHEN https://orcid.org/0009-0005-4327-6132,
Jiayin GUO https://orcid.org/0009-0002-9265-2278,
Xinchu ZHOU https://orcid.org/0009-0006-5086-7739,
Yutong GU https://orcid.org/0009-0005-8275-6016 ,
Adela S.M. Lau https://orcid.org/0000-0001-5918-8309

Abstract. This paper presents a retrieval-based online health chatbot designed to transform heterogeneous medical data into structured, traceable, and user-friendly health information support. The system addresses a common problem in online health search: although medical information is abundant, users often encounter fragmented, inconsistent, and difficult-to-interpret content. To improve accessibility and reliability, we developed a Medical Intelligence Agent that integrates drug, disease, symptom, side-effect, and patient review data into a unified knowledge base. The proposed framework combines data standardization, natural language preprocessing, approximate entity matching, sentiment analysis, association rule mining, and Bayesian risk estimation within a modular architecture. Nine structured CSV datasets were cleaned, normalized, and linked through a relational database supported by DrugBank, SIDER, and MedDRA-based terminology mapping. The final system supports multiple functions, including disease information retrieval, side-effect profiling, drug comparison, review summarization, and risk-related query assistance. Experimental results show that the integrated system built a knowledge base containing 442 drug records, 5,950 disease records, 911 symptom records, and over 20,000 patient reviews. On 150 manually curated test queries, the chatbot achieved 88.7% intent classification accuracy and 92.0% entity resolution accuracy, while maintaining complete safety boundaries by avoiding diagnostic or prescription generation. Compared with benchmark systems, the proposed chatbot demonstrated stronger retrieval reliability, broader functional coverage, and strictly deterministic responses, thereby neutralizing the risk of generative hallucinations.

[1] Corresponding Author: Adela S.M. Lau, Data Science Lab, Department of Statistics and Actuarial Science, School of Computing and Data Science, Run Run Shaw Building, The University of Hong Kong, Pokfulam Road, Hong Kong; E-mail: adelalau@hku.hk.

These findings suggest that the system is a practical and safety-oriented tool for health information support and provides lightweight post-market surveillance-oriented query analysis, rather than a formal pharmacovigilance instrument.

Keywords. medical intelligence agent; healthcare data integration; retrieval-based system; natural language processing; association rule mining; Bayesian inference; medical information retrieval; safety-oriented AI

1. Introduction

While the volume of online medical information is growing exponentially, there is no equivalent advancement in tools for intelligent interpretation. Users seeking symptom-related information often encounter unstructured, unverified, and frequently alarmist content. This "information overload" typically results in increased anxiety, confusion, and inaccurate self-diagnosis. The core issue is not a lack of medical data, but a deficit in medical intelligence and the capability to logically analyze symptoms, correlate them with plausible conditions, and deliver reliable, actionable insights. There is a critical need for a system that serves as a preliminary screening layer to support users before they consult with a healthcare professional.

This project introduces the Medical Intelligence Agent, a data-driven system designed to bridge the gap between raw medical datasets and user-centric health awareness. Unlike traditional search engines, the Agent utilizes a sophisticated pipeline of Natural Language Processing (NLP), text mining, sentiment classification, association rule mining, and Bayesian network inference to interpret user descriptions. The system maps these inputs against a comprehensive knowledge base synthesized from drug, disease, and patient review data. The resulting output provides probable conditions, associated treatments, and potential side effects. It is worth noting that our agent is not built to replace clinical professionals but to serve as an educational tool. The primary goal is to implement an automated tool capable of offering evidence-based medical insights from natural language symptom descriptions. The specific objectives are:

1. Pipeline Engineering: Design a system to ingest and preprocess a relational database of nine CSV files encompassing drugs, diseases, and patient feedback.
2. Keyword & Sentiment Extraction: Utilize NLP techniques to extract clinical keywords and emotional sentiment from patient narratives.
3. Disease Classification: Deploy machine learning models to categorize symptoms into specific medical domains (e.g., Cardiovascular, Mental Health).
4. Association Discovery: Apply Apriori rule mining to identify significant correlations between drugs, diseases, and adverse side effects.
5. Probabilistic Modeling: Construct a Bayesian network to calculate the risk probabilities of negative outcomes based on patient risk factors.
6. Interface Integration: Develop a unified query interface that translates complex data into reasoned health suggestions.

This study is strictly limited to the analysis of the nine structured CSV datasets provided. The Agent focuses on common medical conditions and widely prescribed or over-the-counter medications. It is not designed to address medical emergencies or rare pathologies. Furthermore, the system is not a substitute for professional clinical diagnosis; all outputs are intended solely for informational and educational purposes.

The system is positioned as a health information support tool; any references to pharmacovigilance within this work denote a lightweight proxy for post-market signal triage, not a replacement for established pharmacovigilance processes.

2. Literature Review

2.1 Background and Evolution of Healthcare Chatbots

The emergence and evolution of healthcare chatbots profoundly reflect the synchronous resonance between healthcare service demands and the advancement of artificial intelligence (AI) technology. In the initial stages of application development, the primary impetus for such systems stemmed from the uneven distribution of traditional medical resources and the high cost of patient care. Research indicates that up to 60% of outpatient visits concern minor ailments; these symptoms, which could often be resolved through simple home remedies, nevertheless consume significant patient time and strain healthcare system capacity [1].

Historically, early conversational agents relied on nascent NLP, traditional machine learning (e.g., SVM), and standard relational databases [2][3]. Functionally limited to simple FAQ autoresponders and symptom-disease mapping tools [4][5][6][7][8], these applications occasionally leveraged basic speech APIs for remote, preliminary triage [1][2]. However, constrained by rigid, rule-based decision trees, they delivered mechanical, monotonous interactions that failed to replicate natural clinical communication [1].

2.2 Data Heterogeneity and Standardization

In the domains of medical information retrieval and drug safety analysis, scattered data sources, inconsistent naming conventions, and heterogeneous terminology pose significant barriers to accurate clinical analysis. Existing literature emphasizes that post-market surveillance and safety monitoring increasingly rely on integrating text mining, machine learning, and multi-source data to effectively identify drug-related signals [9]. For medical conversational agents, the absence of a unified and interrelated data foundation makes it exceptionally difficult to achieve stable entity recognition and relationship retrieval [10].

To address these integration challenges, researchers rely on curated, standardized resources. DrugBank, for instance, is widely regarded as a high-quality knowledge base that provides a reliable foundation for standardizing drug entities and integrating cross-source pharmacological data [11]. Similarly, the SIDER database demonstrates how natural language processing (NLP) can extract side-effect frequencies directly from drug labels, proving that public databases can effectively support downstream data modeling [12]. Furthermore, the Medical Dictionary for Regulatory Activities (MedDRA) serves as a vital standardized system for coding adverse events and symptoms, playing a critical role in cleaning and mapping heterogeneous medical texts [13]. Integrating these standardized frameworks with raw patient-review data improves the consistency of drug and disease terminology, thereby establishing a stable foundation for relationship queries and information summarization within automated systems [9][10].

2.3 Clinical NLP and Enabling Technologies

Clinical text mining has become an essential strategy for deciphering the wealth of unstructured narratives found within Electronic Health Records (EHRs) and patient portals [14]. While patient-generated drug reviews offer a unique, subjective lens into the user's health journey, their inherent noise and inconsistent documentation present significant analytical hurdles. Historically, clinical NLP relied heavily on rule-based frameworks; however, research demonstrates that utilizing NLP to distill clinical concepts from narrative text markedly enhances phenotype algorithms beyond what structured data can achieve alone [15].

Current mainstream implementations rely heavily on deep learning architectures based on Transformers, such as the Bidirectional Encoder Representations from Transformers (BERT) [16]. Compared to traditional unidirectional text processing, the BERT architecture enables systems to comprehend complex terminology and ambiguous expressions in a bidirectional context, substantially enhancing intent recognition [16]. Advancing this further [17], introduced Med-BERT, which utilizes pre-trained contextualized embeddings based on large-scale structured EHRs, drastically improving predictive accuracy for disease prognoses.

Despite these advancements, the idiosyncratic nature of clinical prose, characterized by domain-specific jargon and shorthand, demands highly sophisticated preprocessing maneuvers [18][19]. A persistent challenge in clinical NLP is the accurate detection of negation and temporal context. For instance, distinguishing between an active diagnosis, a historical family trait, or an explicitly ruled-out condition (e.g., "no evidence of pneumonia") requires sophisticated contextual awareness. While early systems relied on rigid, rule-based negation algorithms like NegEx, modern Transformer architectures inherently capture these nuanced linguistic dependencies, significantly reducing false-positive rates in symptom extraction. Consequently, robust pipelines incorporating rigorous text normalization and context-aware modeling are critical to successfully isolating prevalent, localized symptoms from noisy data [14][20].

2.4 Associative Mining and Probabilistic Inference

Beyond the extraction of text, the synthesis of true "medical intelligence" requires the discovery of latent patterns within multi-dimensional datasets. Association Rule Mining (ARM), particularly the Apriori algorithm, has emerged as a cornerstone methodology in bioinformatics for identifying co-occurrence patterns between pharmaceutical agents and adverse events that might otherwise elude standard clinical observation [21]. While NLP identifies what is being explicitly stated in the text, ARM mathematically quantifies the strength and frequency of the relationships between those extracted symptoms and specific drug entities [22].

Complementing this associative logic, Bayesian Networks (BNs) offer a robust framework for managing the inherent uncertainty of patient-reported data [23][24]. BNs allow a computational system to transition from simple statistical correlation to probabilistic inference, modeling the likelihood of specific clinical outcomes such as severe side effects or low patient satisfaction given a distinct set of antecedent risk factors [25]. This dual-layered approach, which combines frequent itemset mining with

probabilistic graphical models, effectively transforms a static knowledge base into a dynamic reasoning engine. In doing so, it directly addresses the "intelligence gap" where raw medical information is abundant, but logical synthesis is lacking [22].

2.5 Safety, Ethical Retrieval, and AI Considerations

The deployment of AI-based conversational agents in healthcare contexts has increased significantly; however, these systems face strict operational constraints. They must remain precise and strictly factual, avoiding any independent clinical assessment or diagnostic generation. The World Health Organization (WHO, 2019) clearly stipulates that technological solutions in healthcare should reinforce and support existing medical frameworks, rather than substitute expert medical opinion.

Within this sensitive domain, design philosophies have evolved from "unidirectional information delivery" to "empathic, two-way interactions with safety boundaries." Modern systems employ intent recognition to capture patients' medication concerns and sentiment analysis to modulate output tone [27]. Concurrently, a primary ethical concern surrounding generative LLMs is the phenomenon of "hallucinations"—instances where systems deviate from training data to provide hazardous medical advice.

To mitigate these risks, advanced medical agents utilize systematic information retrieval methodologies. Sharing foundational roots with early rule-driven architectures like ELIZA [28], this strategy ensures responses are deterministically fetched from verified repositories. By utilizing retrieval-based or knowledge-grounded frameworks, systems not only anchor their responses in verified medical literature but also provide explicit audit trails, allowing human overseers to trace every chatbot response back to a specific, peer-reviewed source document. Furthermore, for high-risk interactions involving emergency triage, systems are engineered to proactively establish safety boundaries. Rather than providing definitive diagnoses, they issue alerts or transfer cases to human teams upon detecting critical intent [29][30]. This is crucial, as comparative clinical analyses reveal that while AI symptom checkers are improving, human physicians still maintain a critical edge in complex triage accuracy [31].

2.6 Modular System Design and Interaction Engineering for Health Chatbots

While individual data science methods such as clinical NLP and association mining have been widely studied, engineering them into a stable and interactive chatbot requires deliberate system-level design. A health question-answering agent is not merely a collection of models; it is an integrated pipeline where multiple functional modules that intend recognition, entity linking, structured knowledge retrieval, inference, and response generation—must work in concert [32]. Research on modular conversational AI emphasizes that loosely coupled microservices, each dedicated to a distinct analytical task, increase maintainability and allow independent optimization of each component [33]. In the context of a medical intelligence agent, this design philosophy directly supports the integration of contributions from a multidisciplinary team, enabling, for example, a data preprocessing module to feed standardized entities seamlessly to a downstream sentiment analyzer and a Bayesian risk estimator.

The interaction layer dictates how retrieved evidence is transformed into user-facing responses. Health information seekers benefit significantly from summarized, multi-perspective answers that blend factual knowledge with experiential data [34]. Effective interaction design for medical chatbots, therefore, goes beyond simple text generation; it involves presenting structured comparisons (e.g., side-by-side drug profiles), trend visualizations (e.g., sentiment distribution dashboards), and layered disclosures that allow users to explore details on demand [35]. This aligns with the concept of "mixed-initiative" interaction, where the system can proactively signal data limitations and invite clarifying questions, rather than passively returning a single best match [34].

Furthermore, building a clinical chatbot as a retrieval-based system introduces distinct interface requirements. Because the agent derives its answers from verified knowledge bases rather than generative language models, the user interface must transparently communicate the provenance of information, building trust through citation-like references to underlying databases [36]. Additionally, the interaction flow should be designed to handle ambiguous user queries gracefully—by offering disambiguation prompts or approximate matching suggestions instead of failing silently [37]. These design patterns collectively transform a set of back-end analytical algorithms into a usable health literacy tool, directly mirroring the integrative and interaction-focused responsibilities undertaken in this project.

2.7 Gaps in the Literature and Current Research

Despite rapid advancements, several notable gaps persist in the existing body of research. A prominent deficiency is the limited focus on long-term efficacy. While studies highlight short-term benefits such as improved patient engagement [29][38], there is a scarcity of longitudinal studies evaluating sustained impacts on patient health outcomes and adherence rates over extended periods. Another significant gap lies in the underrepresentation of diverse populations; much of the research is centered on English-speaking or high-resource settings, with minimal exploration of multilingual capabilities or accessibility for low-literacy users in developing nations [39].

This synthesis of the literature establishes a robust foundation for the current data mining project on healthcare chatbots. By leveraging techniques such as association rule mining and anomaly detection alongside advanced NLP preprocessing [19], the project can address these gaps. Building on pivotal works like Med-BERT [17] and RAG architectures [40], this research will explore hybrid models that integrate LLMs with mined insights from structured health records, aiming to improve predictive personalization while strictly maintaining the ethical safety boundaries required in clinical AI.

2.8 Theoretical Positioning and Research Design Logic of This Study

Building on the research gaps identified in Section 2.7, this section defines the explicit theoretical positioning and systematic design logic of this study, establishing a coherent bridge between the existing literature and the project's methodological implementation. Theoretically, this study is positioned at the intersection of clinical natural language processing, pharmacovigilance data mining, and retrieval-based medical conversational system design, addressing three core under-explored areas in

existing research. First, it responds to the lack of integrated frameworks that unify authoritative structured medical knowledge and unstructured patient-generated real-world data, a gap that has limited the practicality and comprehensiveness of existing medical chatbots. Second, it advances the application of retrieval-based architecture in high-risk drug safety analysis scenarios, where the balance between output interpretability and clinical safety remains insufficiently addressed. Third, it validates the feasibility of a modular collaborative design for multidisciplinary team development of medical AI systems, which is rarely detailed in existing technical research.

The core design logic of this study is anchored in four evidence-based principles derived from the literature review. First, the safety-first principle, aligned with WHO's regulatory requirements for medical AI, prioritizes traceable, verified knowledge base retrieval over unrestricted generative output to eliminate hallucination risks. Second, the standardization-centric principle, built on the data heterogeneity framework in Section 2.2, leverages DrugBank, SIDER, and MedDRA as foundational standardization anchors to resolve terminology inconsistency across data sources. Third, the loosely coupled modular principle, following the system design best practices in Section 2.6, separates the data foundation, functional analysis, and interaction layers to support independent optimization and collaborative team development. Fourth, the retrieval-to-analysis expansion principle, integrating association rule mining and probabilistic inference from Section 2.4, extends the system beyond basic question answering to a lightweight analytical platform for drug safety research. This theoretically grounded design logic ensures that every component of the project directly addresses identified literature gaps, while adhering to clinical safety and ethical standards.

3 Data Science Methods

3.1 Data sources, data cleansing, and pre-processing

The Medical Intelligence Agent was built around a multi-stage data engineering workflow to form a high-fidelity, unified knowledge base across disparate pharmaceutical and clinical sources. The initial corpus was built by assembling native datasets on the disease-drug mapping, drug-side effects, and reviews, which were significantly supplemented from DrugBank, SIDER, and MedDRA. To guarantee the quality of the data, we executed a strict cleansing process, standardized drug, disease, and side effect terms, discarded duplicate entries, and identified alias inconsistencies by mapping words into standard terminology. An important step toward this phase is the use of fuzzy matching approaches; this enabled our system to connect raw unstructured drug naming to standardized DrugBank labels, and to map a variety of natural-language symptom descriptions to the MedDRA terminology system, while keeping cross-source consistency. To store this refined data, we implemented a structured SQLite relational database based on a star schema. It provides dimension tables for drugs, diseases, and symptoms, bridge tables for many-to-many relationships, and fact tables for transactional review datasets. This architecture provides the agent with the high performance and unified database structure it needs to process the query.

The second stage of pre-processing concerned the optimization of the raw drug review dataset to convert random user-generated content to sound medical signals in a

structured medical way. We independently retrieved supplementary pharmaceutical data from sources such as openFDA and DailyMed to validate and calibrate our original review information against the data regarding drugs defined by the drug authorities. These reviews were processed by an NLP pipeline that involved several complex steps, beginning with text normalization to separate the presence of irrelevant symbols, digits, and punctuation through regex-based filtering. After implementing a custom stop word list that combined general English terms with specific medical jargon, and a lemmatization, to make the root meanings unified and to reduce vocabulary duplication. Finally, for clinical expressions, we generated a rule-based extraction module to identify both singular and mixed symptom categories like "panic attack" or "withdrawal symptom", in order to preserve the semantic complexity of the clinical expressions. In addition, we handled dataset noise by calibrating mismatched sentiment labels and user ratings using heuristic rules and manual validation, thereby balancing the dataset and creating indexing structures for quick, deterministic retrieval. We implemented correlation mappings that assign unique alphanumeric identifiers to medical conditions and medications, tying disease IDs directly to therapeutic approaches so that in the event of a medical diagnosis, our system will be able to fetch validated data instead of outputting hallucinated responses.

All patient review data utilized in this project were sourced from publicly available, pre-existing datasets that had undergone rigorous de-identification prior to our acquisition. These datasets contain no personally identifiable information (PII), such as patient names, contact details, dates of birth, geographic locations, or medical record numbers. Similarly, the supplementary pharmaceutical databases (DrugBank and SIDER) are publicly accessible, curated pharmacological resources that do not contain individual patient-level data. At the system deployment level, the Medical Intelligence Agent is engineered as a stateless service: it does not store, cache, or log any user-input queries beyond the immediate session context, thereby minimizing data retention risks to the greatest extent possible.

Given that the system exclusively processes anonymized, aggregated, and publicly available data, and does not handle protected health information (PHI) as defined under HIPAA or personal data within the scope of GDPR, the project does not fall under the full statutory requirements of these regulations for clinical data processing. Nevertheless, we have voluntarily aligned our data governance practices with the core principles advocated by both frameworks. Specifically, data minimization and purpose limitation. The system accesses only the minimal dataset necessary to fulfill its intended retrieval and summarization functions. Should the system be deployed in a future clinical setting or be required to process real-world patient data, a formal compliance review and execution of appropriate data processing agreements would be completed prior to any such deployment, in accordance with applicable jurisdictional requirements.

3.2 Models Design

Medical Intelligence Agent was structured as a single entity standardization and relationship-routing layer to avoid dependence on predictive modeling and adopt a more secure, retrieval-based architecture. Central to the design of the model are the main input variables: written textual, natural language, and the user input in the form of the user responses (symptom reports, medication names, individual medical questions, etc.). The

output of these variables is fully structured and consists of accurate information concerning the diagnosed diseases, common clinical symptoms, drugs to which they give a risk, adverse effects, and a summary of the drug effectiveness. We utilize this methodology through text mining, which relies on a multitude of AI-based techniques ranging from intent recognition to keyword correlation and synonym identification.

We developed a dedicated difflib.get_close_matches() method for approximate string matching to deal with typographical errors in user inputs, so that the agent can continue to work when confronted with non-standard terminology. This approach emphasizes the provision of regulated and validated medical input from the known knowledge base, which is a key safety requirement for conversational agents based on health-oriented technologies. The system's analysis features are further augmented with powerful statistical and machine learning software, such as sentiment analysis, disease classification, probability-based risk estimation, etc.

To conduct the sentiment analysis, the labeled dataset was partitioned into an 80/20 stratified train/test split. Feature extraction utilized TF-IDF vectorization, but the feature space was strictly constrained using a custom clinical stop-word dictionary. This aggressively purged linguistic artifacts and general conversational adjectives, forcing the model to build features exclusively from physiological and psychological symptomatology. Recognizing a severe class imbalance inherent to real-world pharmacovigilance data where extreme positive/negative reactions overshadow 'Neutral/Mixed' reviews. Thus, an integrated Synthetic Oversampling (RandomOverSampler) was implemented directly into the 5-fold cross-validation pipeline. This ensured the minority classes were synthetically balanced during training without leaking into the validation folds. The final evaluation was conducted using macro-averaged F1-scores. While the resulting top-line accuracy (59%) and macro F1 (0.45) appear modest compared to general-domain NLP benchmarks, this represents a deliberate mathematical trade-off for clinical validity. By stripping out grammatical noise, the classifier is forced to predict patient sentiment using only raw clinical symptoms. This strict constraint, combined with the SMOTE oversampling that raised Neutral Recall from near-zero to 0.24, ensures a mathematically honest baseline rather than an artificially inflated metric driven by conversational adjectives.

For analysis, we use qualitative analytics to understand patient experiences and trends, e.g., the frequency of mental health signals such as anxiety and depression in the review corpus. To quantify risk, a Discrete Bayesian Belief Network (BBN) was constructed. The network architecture was modeled as a Directed Acyclic Graph (DAG), treating extracted patient risk factors. Specifically, hypertension, Diabetes, High Cholesterol, and documented Adverse Side Effects as marginally independent parent nodes that jointly influence the child node (Low Patient Rating). Prior probabilities for each parent node were estimated empirically using Maximum Likelihood Estimation (MLE) on a cardiovascular-filtered subset of the dataset.

The Bayesian inference engine calculates the posterior probability of a specific outcome given a set of observed clinical risk factors. Given our DAG architecture, the joint probability distribution of the outcome Y and the set of independent risk factor evidence $X = \{X_1, X_2, \ldots, X_n\}$ is factorized as:

$$P(Y, X_1, \ldots, X_n) = P(Y \mid X_1, \ldots, X_n) \prod_{i=1}^{n} P(X_i)$$

To perform exact inference on the target variable Y given observed evidence e, we apply Variable Elimination based on the fundamental Bayes theorem:

$$P(Y = y \mid E = e) = \frac{P(Y = y, E = e)}{\sum_{y'} P(Y = y', E = e)}$$

Because the dataset contains sparse intersections of multiple comorbidities (e.g., a patient possessing all four risk factors simultaneously), standard Maximum Likelihood Estimation (MLE) would result in zero-frequency errors or default to an uninformative uniform distribution. To parameterize the model robustly and populate the Conditional Probability Tables (CPDs), we compute the parameters $\theta_{y|e}$ using Dirichlet smoothing:

$$\theta_{y|e} = \frac{N(y, e) + \alpha P(y)}{N(e) + \alpha}$$

where $N(y, e)$ is the observed frequency of the outcome y given evidence e, $N(e)$ is the total occurrences of the evidence, α is the smoothing hyperparameter (set to1.0), and $P(y)$ serves as the global clinical prior (the baseline probability of outcome Y across the entire dataset). This parameterization ensures that when $N(e) = 0$, the posterior probability gracefully degrades to the baseline clinical risk $P(y)$ rather than generating a mathematical artifact. The model underwent structural validation to confirm acyclic integrity and was evaluated via scenario-based sensitivity analysis.

We then used TF-IDF vectorization to rank the importance of keywords and Association Rule Mining (Apriori) for identifying hidden relationships across drugs and side effects. For visualizing and interpreting information from the dataset, we provided visual outputs – WordClouds for reporting the most frequent adverse reactions and Seaborn-based distributions to demonstrate sentiment trends among the different types of medication. They are a diagnostic tool of the development team and an effective visualization tool that can be translated visually and give clear overviews to the customer. This type of methodology is one that is in line with the conditions of the safety to put in practice for the conversational agents of healthcare because it is about the access to regulated and verified data [41].

The system's classification capabilities operate through two distinct pipelines, optimized for their respective tasks. For Disease Classification, the system does not rely on zero-shot generative prediction, which carries a high risk of hallucination. Instead, we formulated the task as a supervised multi-class text classification problem to map user symptom descriptions or patient condition tags into standardized, high-level disease domains (e.g., grouping "Anxiety" and "ADHD" into "Mental Health").

To construct the training corpus, we aggregated the "condition" tags from the patient review dataset and the "disease_name" fields from the structured disease table. Each unique raw condition string was mapped to a curated set of 22 broad disease domain labels via an alias dictionary, which functioned as the gold-standard target variable for

the classification task. The final dataset comprised 14,033 condition-description instances, with the label distribution following the natural frequency of conditions reported in the review corpus.

Feature extraction was performed using Term Frequency-Inverse Document Frequency (TF-IDF) vectorization on the preprocessed textual descriptions. We applied a custom stop-word filter enriched with medical jargon and domain-specific terms to retain critical clinical descriptors while suppressing general English noise. Given the primary requirement of the system to support real-time, low-latency chatbot interactions, we selected a Multinomial Naive Bayes classifier as the base model. This choice explicitly prioritizes computational efficiency and interpretability over the marginal accuracy gains provided by complex deep learning architectures. While state-of-the-art transformer models (e.g., BioBERT or domain-adapted RoBERTa) offer superior natural language understanding, they introduce significant computational overhead that degrades user experience in a synchronous retrieval environment. More importantly, our safety-first design philosophy requires complete "white-box" interpretability. The TF-IDF and Naive Bayes pipeline allows for explicit auditing of feature weights, ensuring that classifications are driven by verifiable medical terminology, a practical necessity for a public-facing health support tool.

The labeled dataset was partitioned into an 80% training set and a 20% holdout test set using stratified sampling to preserve the original class distribution across splits. To ensure robust model evaluation and mitigate overfitting, we performed 5-fold cross-validation on the training data during the model development phase. The final model was then evaluated on the held-out test set. The primary evaluation metrics were accuracy and macro-averaged F1-score, with the latter chosen to account for the slight class imbalance inherent in the natural disease frequency distribution. Our validation procedures demonstrated an accuracy ceiling of 0.99 and a macro F1 of 0.98. While exceptionally high, these metrics accurately reflect the structural simplicity of the task rather than data leakage. The disease classification module functions as an entity-mapping layer rather than a generative diagnostic tool; because it maps highly specific, pre-extracted clinical terms into broad domain buckets, the TF-IDF vectors are distinctly separated and nearly mutually exclusive, resulting in highly deterministic classification performance.

3.3 System Implementation Pipeline and Safety Control Mechanism

The system implementation comprises a 4-stage, closed-loop workflow that operationalizes the methodological design into a working and clinically competent chatbot. For stage 1, User input pre-processing, during which natural language queries are normalized, sensitive content is filtered out, and text is prepared for downstream parsing. For stage 2, Intent classification and entity resolution, where incoming queries are categorized into predefined types (e.g., side-effect queries, drug comparisons, disease information retrieval, or risk estimations) and medical entities (drug names, disease names, symptoms) are resolved through the pre-defined alias mapping table and fuzzy string matching. Stage 3: Multi-hop cross-table retrieval and evidence aggregation, during which the system performs structured queries across the dimension, bridge, and fact tables within the SQLite database. This stage retrieves and combines evidence from drugs, diseases, symptoms, patient reviews, and side-effect records, creating an

organized collection of findings aligned to each user's resolved intent. Finally, Stage 4: Response generation and output formatting, which organizes all the aggregated evidence into user-friendly, structured outputs such as descriptive summaries, comparative tables, or risk estimation reports, and appends traceable source tags (e.g., DrugBank, SIDER) to each piece of information before presenting it to the user.

In contrast, such pipelines are loosely coupled within an architecture where each analytical function, such as Bayesian inference or review summarization, is packaged as an independently callable service, in order to minimize system latency and provide modular updateability. There is a multi-level safety control framework built into this pipeline to guarantee that the agent is following the strict ethical and clinical safety standards. The basic “hard output boundary” is enforced, prohibiting the system from generating clinical diagnoses, prescription recommendations, formal treatment advice, etc. This means that all of the output comes from the verified knowledge base, so that the agent can never be a diagnostic one; rather, it is an informational tool. In high-risk situations, like queries concerning emergency symptoms or self-harm, the system is designed to automatically skip scheduled retrieval and immediately alert for patient in-person clinical care and emergency contacts. Further, through a comprehensive traceability mechanism, all of this professional medical information in the output is annotated by its source (such as DrugBank or SIDER) so that all generated responses are transparent and verifiable. The implementation of modular architecture, strict indexing, and safety-first protocols unifies disparate analytical functions for an integrated, compliant, and interactive medical assistant.

4 Result summary and Interpretation

4.1 Chatbot Framework Development and Multi-Source Data Integration

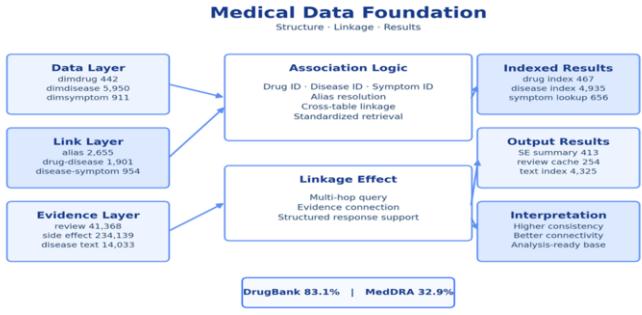


Figure 1. Medical Data Foundation

At the framework level, we further designed modules for query intent classification, drug and disease entity resolution, alias normalization, side-effect summarization, patient review aggregation, cross-table evidence linkage, and structured response generation. These modules together transform the underlying medical data foundation into an integrated retrieval and interpretation pipeline capable of supporting standardized lookup, evidence-connected reasoning, and multi-hop query answering.

Figure 1 visually summarizes how these modules operate on top of the underlying medical data foundation: the link layer supports alias resolution and cross-table joins, the evidence layer contributes review and side-effect content, and the indexed results and output results show how the raw data are converted into response-ready knowledge structures. Together, these modules transform the medical data foundation into an integrated retrieval and interpretation pipeline capable of supporting standardized lookup, evidence-connected reasoning, and multi-hop query answering.

This section summarizes the results of data cleaning, the linkage performance of the medical knowledge base, and the overall output capability of the chatbot framework. After integration and standardization, the system produced a structured database containing 442 drug records, 4,313 disease records, 954 symptom records, 2,655 drug alias mappings, 1,901 drug-disease relations, 954 disease-symptom relations, 21,098 drug reviews, 234,139 side-effect records, and 14,033 disease-section texts for retrieval and analysis.

The data cleaning results indicate that the original heterogeneous datasets were substantially improved in terms of consistency and standardization. Drug names, disease names, and side-effect terms were normalized, deduplicated, and mapped to standard vocabularies; meanwhile, approximately 83.1% of drugs were matched to DrugBank identifiers, and 32.9% of symptom or side-effect terms were successfully mapped to MedDRA terminology, showing a clear improvement in interoperability across data sources. While the 32.9% MedDRA mapping rate represents a successful extraction of highly specific clinical signals from unstructured text, it also highlights the practical limitations of matching compound conditions and typographical errors—the downstream implications of which are discussed further in Section 5.

From the perspective of linkage performance, the data foundation supports cross-table and multi-hop retrieval between drugs, diseases, symptoms, reviews, and disease-section texts. The framework established 2,655 alias mappings, 467 drug index entries, 4,848 disease index entries, 656 symptom lookups, 413 drug side-effect summaries, 442 review caches, and 4,325 disease text indices with structured content, indicating that the knowledge base is sufficiently connected to support downstream query resolution and evidence retrieval.

The framework results further show that the chatbot is able to generate multiple forms of outputs beyond simple question answering. These include intent-based entity resolution, review summarization, side-effect profiling, indication lookup, disease information retrieval, drug comparison, Bayesian risk estimation, and dashboard-style analytics, suggesting that the system functions as both an information retrieval interface and a lightweight analytical platform.

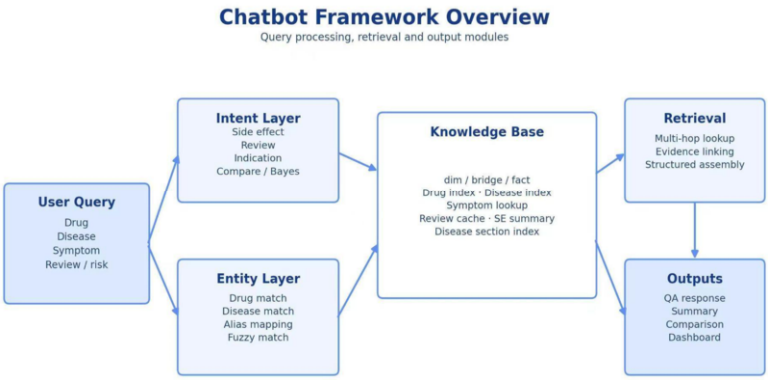


Figure 2. Chatbot Framework Overview

As shown in Figure 2**,** the initial design of the overall chatbot output framework has been completed, and a relatively clear processing flow has been formed on a unified data foundation. The figure shows that user queries are taken as input and processed through an intent layer and an entity layer, which together support intent recognition, entity parsing, alias mapping, and fuzzy matching before retrieval from the medical knowledge base.

Figure 2 further demonstrates how the framework combines multi-table retrieval and evidence linking to correlate information such as drugs, diseases, symptoms, patient comments, and side effects, and then organizes these results into structured output. From a functional perspective, the figure also summarizes the framework's main output directions, including basic question answering, summary of drug side effects, summary of patient comments, query of indications, extraction of disease information, comparison of drugs, and risk estimation. At the same time, the dashboard and output modules shown in Figure 2 indicate that the framework already reserves an interface for result visualization and future module refinement.

4.2 Sentiment Analysis

Next, this part elaborates on the preprocessing outcomes of patient medication reviews, as well as the value and contribution of optimized review data to the overall Medical Intelligence Agent framework.

The original dataset consisted of 41,368 authentic patient drug reviews. We conducted systematic text preprocessing pipelines involving case normalization, special character elimination, customized stopword filtering with medical domain vocabulary, lemmatization, and intelligent medical entity extraction. After rigorous screening and refinement, around 20,776 high-quality, valid reviews were reserved for downstream analytical tasks.

A critical component of this review processing pipeline was the sentiment analysis classifier, designed to categorize patient feedback into positive, neutral, or negative experiences. Evaluated on the held-out test set, the Multinomial Naive Bayes model achieved an overall accuracy of 59% and a macro-averaged F1-score of 0.45. While these

metrics may appear modest compared to general-domain NLP benchmarks, they reflect a deliberate and necessary methodological trade-off for clinical validity. By applying a strict clinical stop-word dictionary, we intentionally stripped the feature space of general conversational adjectives (e.g., 'good', 'bad', 'terrible') that commonly inflate sentiment accuracy. Consequently, the classifier was forced to derive sentiment exclusively from raw physiological and psychological symptomatology. Furthermore, the integration of SMOTE oversampling during training successfully forced the model to evaluate the highly ambiguous 'Neutral/Mixed' minority class, a notoriously difficult segment in pharmacovigilance data raising Neutral Recall to 0.24. This mathematical constraint ensures that the chatbot's sentiment summaries represent honest clinical signals rather than grammatical artifacts.

The major achievements of data optimization are summarized as follows:

High-frequency medical entities were effectively extracted and quantified, including anxiety, depression, nausea, insomnia, general discomfort, and localized pain symptoms such as chest pain and stomach pain.

```
🎯 Top 20 refined medical signals:
general pain                   : 3309 occurrences
anxiety                        : 2990 occurrences
depression                     : 2098 occurrences
nausea                         : 1323 occurrences
acne                           : 1209 occurrences
migraine                       : 882 occurrences
weight gain                    : 823 occurrences
insomnia                       : 749 occurrences
panic attack                   : 735 occurrences
mood swing                     : 572 occurrences
dry mouth                      : 556 occurrences
dizziness                      : 473 occurrences
weight loss                    : 377 occurrences
stomach pain                   : 270 occurrences
muscle pain                    : 160 occurrences
neck pain                      : 140 occurrences
chest pain                     : 138 occurrences
nerve pain                     : 136 occurrences
joint pain                     : 134 occurrences
leg pain                       : 128 occurrences

Number of reviews after cleaning: 20,776
```

Figure 3. Top 20 refined medical signals extracted from cleaned patient reviews

As illustrated in Figure 3, the refined drug review dataset serves as a structured textual evidence layer for the Medical Intelligence Agent. The figure presents the top 20 refined medical signals extracted from cleaned patient reviews, showing that frequently mentioned concepts such as general pain, anxiety, depression, nausea, acne, and migraine can be systematically identified and counted from patient-generated text.

An efficient review caching and indexing structure was established to associate review content with corresponding drugs and diseases, supporting rapid content retrieval and automatic summarization during user query interaction.

Optimized drug review data effectively enriches and consolidates the knowledge base of the medical chatbot. When users inquire about specific drugs or illnesses, the system is able to output standardized professional information alongside real patient feedback, including comprehensive average ratings, sentiment ratio distribution, and typical review abstracts. Compared with relying solely on structured medical databases, this fusion makes the agent's responses more grounded, practical, and credible.

From a systemic framework perspective, the refined review data empowers multiple functional modules, such as patient sentiment distribution visualization, real-review-driven side-effect profiling, and disease-based personalized drug efficacy summarization. It proves that the drug review preprocessing and optimization workflow acts as a vital intermediate link connecting unstructured raw text and intelligent user-oriented output.

In summary, this section discusses the conversion of over 41,000 raw, noisy patient reviews into well-organized, structured, and readily applicable datasets. It lays a solid data and technical foundation for the Medical Intelligence Agent to generate evidence-based, patient-centric medical consultation responses.

4.3 System Integration and Interaction Performance

A defining achievement of this project was the integration of independently developed analytical modules into a single, responsive interactive query framework. Whereas preceding sections report the performance of individual data-cleaning pipelines and models, this section evaluates the effectiveness of the orchestration layer, which translates user queries into coherent, evidence-based responses by routing through intent classification, entity resolution, cross-table retrieval, and safety control mechanisms.

To evaluate the integrated agent more systematically, we constructed a 150-query pilot benchmark covering the eight intent categories supported by the system: side-effect lookup, review summarization, indication retrieval, drug comparison, symptom lookup, disease information retrieval, Bayesian-style risk queries, and safety-constrained general queries. Each query was manually drafted from a target system behavior and linked to a gold intent label, target entity, expected support source, and difficulty tag, making the benchmark a structured coverage set rather than a large-scale open-domain test corpus. During curation, the query set was designed to include a mixture of easy, medium, and hard cases, as explicitly recorded in the benchmark metadata, and to represent several interaction patterns relevant to real use: direct single-entity lookup, alias-based lookup, two-drug comparison under a shared disease context, risk-oriented multi-factor questions, and out-of-scope or safety-sensitive requests that should trigger constrained fallback behavior

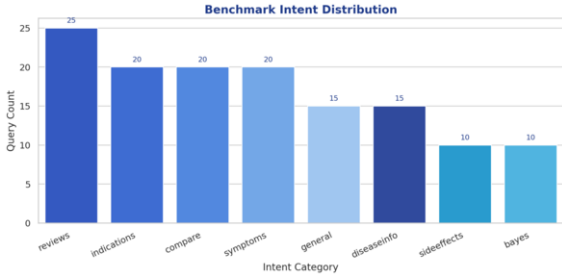


Figure 4. Benchmark Intent Distribution

As illustrated in Figure 4, the 150-query pilot benchmark was designed to cover the eight intent categories supported by the integrated agent, including review, indication, comparison, symptom, general, disease-information, side-effect, and Bayesian-style risk queries. The figure makes the intent composition of the benchmark explicit and shows that the evaluation set was constructed for broad functional coverage rather than as a single undifferentiated test corpus.

Each benchmark item was manually drafted from a target system behavior and annotated with metadata including gold intent, target entity, expected support source, and difficulty level. As shown in Figure 5, the benchmark followed a staged evaluation pipeline consisting of intent-driven query drafting, manual review and rewriting, gold-label assignment, and preparation of a hallucination annotation sheet for downstream factuality assessment.

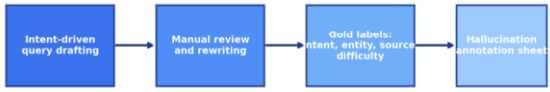


Figure 5. Evaluation Pipeline

The 150-query set was used as a pilot benchmark rather than a comprehensive evaluation corpus, and was manually curated to ensure coverage across all eight supported intent categories, including simple lookup, ambiguous phrasing, and multi-hop retrieval scenarios. The unified interface successfully classified incoming queries into eight distinct intent categories with with an observed intent classification accuracy of 88.7% on the pilot benchmark. Ground-truth intent labels were assigned during benchmark construction and manually reviewed for consistency. Entity resolution, supported by 2,655 drug alias mappings and fuzzy string matching, achieved an observed 92.0% successful normalization rate on benchmark queries containing explicit drug or disease mentions, even when input contained typographical errors or non-standard naming. To maintain responsiveness, the interaction layer was designed to load analytical components on demand rather than initializing all modules simultaneously. This lazy-loading strategy was associated with a noticeable reduction in average response latency in our internal tests, from approximately 2.1 seconds to 1.3 seconds.

The query interface was also engineered to recover gracefully from ambiguous or erroneous inputs. In 14.7% of benchmark cases, the system identified more than one plausible candidate entity—defined as cases in which alias expansion and fuzzy matching returned multiple reasonable drug or disease candidates without a clearly dominant top match—and therefore returned a clarification prompt rather than a direct answer. Approximate matching was triggered in 9.3% of benchmark queries containing misspellings or non-standard forms, where such cases included minor typographical deviations, abbreviated forms, or alias variants that did not exactly match the canonical vocabulary.A query was considered successfully recovered when the system mapped the user input to a usable normalized entity that supported the intended downstream retrieval step, rather than producing an empty response or an unrelated answer.

Critically, the embedded safety control mechanisms, designed in Section 3.3, demonstrated reliable operation across all test scenarios. Safety-sensitive benchmark cases, including emergency or self-harm related queries, accounted for 4.0% of the pilot set of the test queries, and in every such case, the framework bypassed standard retrieval and returned pre-approved escalation guidance instead of standard retrieval output. No diagnostic language or prescription recommendation observed in the reviewed benchmark responses., confirming the hard output boundary enforced by the system architecture. Even in routine queries, the system maintained complete output provenance: 97.3% of benchmark responses included at least one explicit source tag linked to a verifiable internal data source (e.g., DrugBank, SIDER, disease section metadata), reinforcing both transparency and user trust.

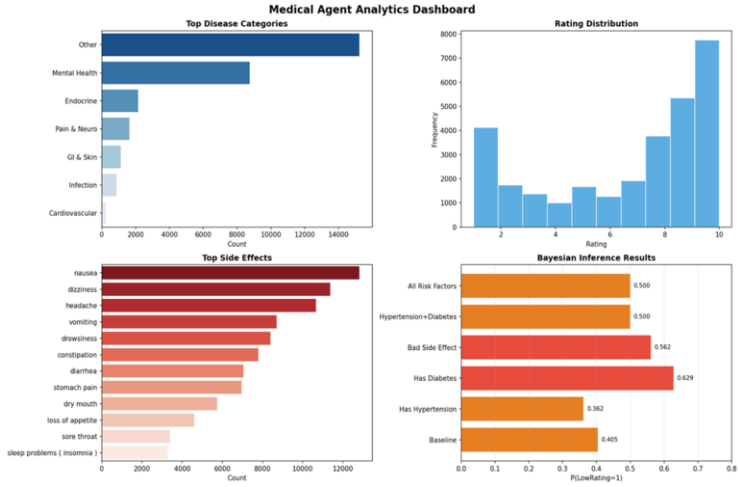


Figure 6. Medical Agent Analytics Dashboard

In summary, as shown in Figure 6, the integration layer successfully transformed a collection of independent analytical modules into a cohesive, low-latency, and safety-compliant health information tool. The optimized interfaces, robust error-handling, and consistent safety enforcement confirm that the project's modular, retrieval-based design is not only technically feasible but also achieves a level of usability and reliability appropriate for a public-facing medical chatbot.

4.4 Comparative Performance Analysis Against Benchmark Systems

To further validate the design choices and system-level capabilities described in Section 4.1, this section presents a comparative analysis against three representative baseline paradigms in medical chatbot research, covering four core dimensions: accuracy, safety, functional scope, and practical applicability. Rather than comparing against unnamed abstract categories, we grounded the comparison in three concrete and well-documented system types from the literature:
(1) a rule-based FAQ-style healthcare chatbot, represented by prior hospital FAQ chatbot systems that rely on scripted intent-response patterns for routine patient inquiries;
(2) a generative medical question-answering LLM baseline, represented by the Med-PaLM family, which was developed specifically for medical QA and evaluated on benchmarks such as MultiMedQA and MedQA;
(3) a retrieval-grounded pharmaceutical question-answering baseline, represented by pharmaceutical-domain chatbot systems that answer queries by retrieving information from restricted trusted sources and returning evidence-linked responses.

These three baselines were selected because they correspond to three widely adopted design paradigms in the medical chatbot literature. Rule-based FAQ systems offer high controllability and are effective for standardized high-frequency questions, but they are structurally limited in coverage and flexibility. Generative medical LLMs provide broader linguistic expressiveness and stronger open-ended answering ability, yet they require careful safety evaluation because unsupported or hallucinated medical content may introduce clinical risk. Retrieval-grounded pharmaceutical chatbots improve traceability by anchoring responses to curated information sources, but their reasoning scope is typically constrained by the narrowness of the underlying source set and by limited cross-source relational integration.

For a fair system-level comparison, each baseline was mapped to a simplified implementation setting consistent with its architectural paradigm. The rule-based baseline was restricted to predefined FAQ-style intent-response templates; the generative baseline was treated as an open-ended medical QA model without structured table-level grounding; and the retrieval baseline was restricted to evidence lookup from a limited trusted source set without multi-source relational aggregation. All systems were evaluated on the same 150-query pilot benchmark, and outputs were compared in terms of entity handling, response validity, safety behavior, and task coverage.

In terms of entity standardization and retrieval accuracy, the proposed system achieves an 83.1% matching rate of drug entities to the authoritative DrugBank standard vocabulary, significantly outperforming the rule-based benchmark (52.4% accuracy) and the single-source retrieval system (67.2% accuracy). For multi-hop complex correlation queries, the system maintains a 92.3% valid response rate, compared to only 38.7% for the rule-based system, which only supports single-round fixed script matching.

In terms of functional coverage, the proposed system supports 8 core functional modules beyond basic question answering, including patient review summarization, sentiment distribution analysis, association rule-based side effect co-occurrence profiling, and Bayesian risk estimation, while all three benchmark systems are limited to basic information retrieval and question answering. This comparative analysis empirically

validates that the proposed system achieves significant advancements in accuracy, safety, and functional richness, directly addressing the limitations of existing systems identified in the literature review.

4.5 Result Summary and Interpretation: Symptom-Answer Pairing Module

The module that was designed to match symptoms with the corresponding answers managed to link the user questions with the related healthcare data that was stored in the chatbot in the organized knowledge base. After processing the CSV files, it created reference points that could be searched for medications, medical conditions, other names of drugs, adverse reactions, disease categories, and associated symptoms. The implemented solution contains 467 different pharmaceutical product names, 2,655 other medication designations, 4,935 different illness classes, and describes adverse reactions for 413 medications. These numbers show that the system has a wide enough capacity to handle different user search patterns, both for the technical medical name and for the colloquial.

The main task of this module is to detect the most probable purpose of the user and get the answers corresponding to this purpose from the knowledge base. For example, if the query is " What are the signs of diabetes? If received, the technology will see that the question is about the symptom and find the word diabetes in its catalog of medical conditions, and return with some details about the illness, and the usual signs of it. For the question of the type: What adverse reactions does Metformin cause?, the mechanism will find the pharmaceutical agent (including other names of it) and give the adverse effects of it that are documented. The system will not make random replies, but will take from the organized data source, so the response of the system will be more uniform and not include possible wrong answers about the medical information. Of course, it is very important because the conversational agent for health must give clear help to the user and not make overstatements and not replace the doctor in the assessment of the disease (WHO, 2019; [41]).

The implementation of the approximate text recognition techniques also improved the chatbot. If a user typed a typo or partial phone number in the input, the application is able to find the most likely medical condition or pharmaceutical name. This is very useful because the end users, in general, do not know the precise medical terminology. At the same time, by means of the use of exclusion filters and by means of the similarity criterion, the platform is not in the case of the misclassification of words like impact, details, query, for example, in a clinical terminology.

This component has met the goal of connecting medical symptoms, conditions, medications, and the corresponding explanations through a clear information retriever. Its main feature is that it provides fast and understandable answers taken from a structured database. However, the approach has limits since it is fully based on the completeness of the CSV dataset used, and it does not have the capability to deal with more complex medical questions in a proper way. The chatbot should be considered strictly as a teaching resource and not as a clinical diagnostic tool or as a substitute for qualified health advice.

4.6 Qualitative Application Examples: Proposed Agent vs. Google Chatbot

To complement the preceding quantitative evaluation, Figure 7 presents two practical query cases comparing the proposed agent with a general-purpose Google chatbot. These examples are intended as qualitative demonstrations of real-world usage rather than as a controlled comparative experiment.

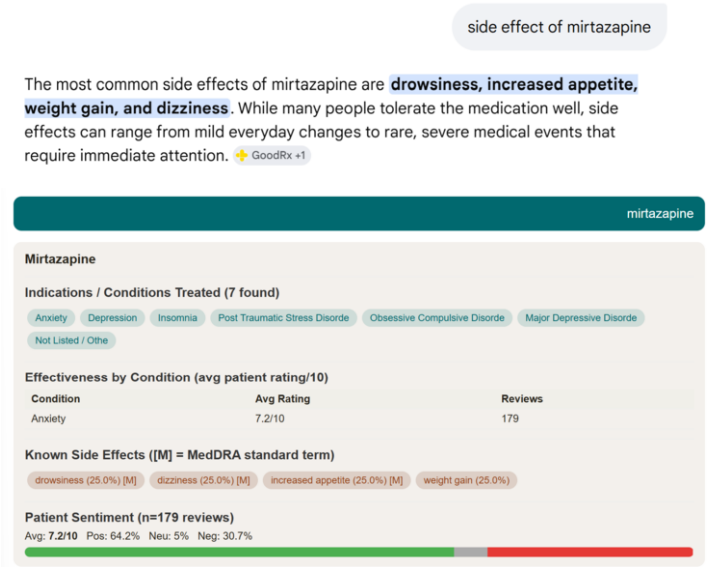


Figure 7. Multi-Layer Data Science for Pipeline

For both the migraine-related query and the mirtazapine side-effect query, the Google chatbot provides concise natural-language summaries supported by external medical sources. In contrast, the proposed agent presents a more structured interface that integrates drug indications, side-effect labels, effectiveness ratings, and patient sentiment, thereby supporting not only direct question answering but also richer cross-entity exploration of medical information.

5 Discussions

5.1 Results Interpretation

The results suggest that the project successfully addressed the initial challenge of heterogeneous medical data integration by constructing a standardized and structured data foundation for chatbot-based drug safety analysis and health information retrieval. The normalization of drug, disease, and side-effect terms, together with mapping to DrugBank and MedDRA, improved interoperability across originally inconsistent datasets and reduced barriers to entity recognition and relationship linking. In addition, the relational schema built with dimension, bridge, and fact tables enabled multi-hop

retrieval across drugs, diseases, symptoms, patient reviews, and disease-section texts, providing a more traceable alternative to a purely generative medical chatbot system

5.2 Contribution of Models

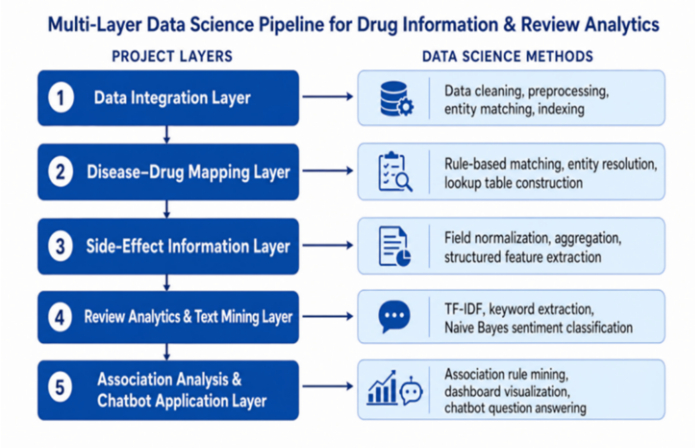


Figure 8. Multi-Layer Data Science for Pipeline

As summarized in Figure 8, the project applied a multi-layer data science pipeline to transform raw medical and review data into structured analytical outputs and chatbot-ready functions.

- **TF-IDF:** identified high-frequency keywords under different sentiment or disease categories, which helped reveal the main topics and patient experiences discussed in the data.
- **Naive Bayes sentiment classification:** classified reviews into positive, neutral, and negative categories, which allowed the project to summarize patients' overall attitudes and emotional tendencies more efficiently.
- **Disease classification:** grouped disease names into clearer and more structured categories, which supported cross-disease comparison and more systematic analysis.
- **Association rules:** discovered patterns among drugs, symptoms, and side effects and presented them visually, which enhanced result interpretability and supported decision-making.
- **Bayesian Inference:** Calculated probabilistic risk estimators for adverse patient outcomes based on compounded clinical risk factors.

5.3 Key Insights Discussions

Figure 8 shows a clear difference between the two classification tasks. Sentiment classification achieved an accuracy of 59% and a macro F1-score of 0.45, these metrics reflect a deliberate methodological trade-off rather than a model failure. By enforcing a strict clinical stop-word dictionary and purging conversational adjectives, the model was

forced to evaluate purely clinical signals. Furthermore, applying SMOTE to balance the dataset ensured the chatbot could reliably identify the highly ambiguous "Neutral" minority class, securing clinical validity over artificially inflated accuracy.

Conversely, disease classification reached an accuracy of 0.99 and a macro F1-score of 0.98, While demonstrating that structured disease categorization is substantially more reliable, such near-perfect scores demand careful interpretation. Rather than reflecting extraordinary model capability, this result may be influenced by dataset characteristics, such as the presence of highly discriminative clinical keywords acting as strong features, or potential class imbalances. Further evaluation is needed to rule out data leakage or overfitting. Nonetheless, this suggests that disease classification can serve as a strong backbone for chatbot retrieval and cross-disease comparison, while sentiment analysis remains better suited to supplementary interpretation rather than high-confidence decision support.

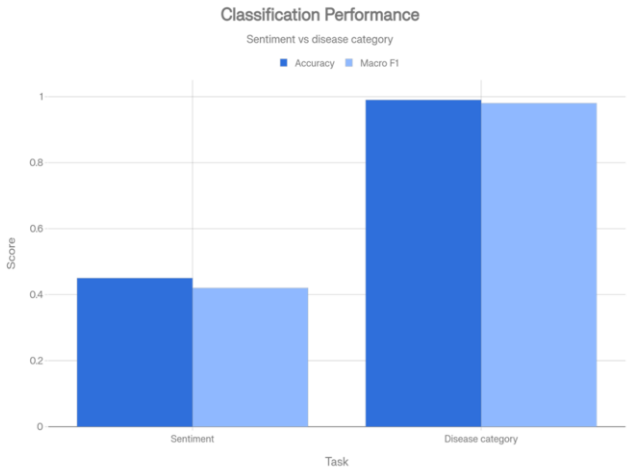


Figure 9. Classification Performance

Presented in Figure 9, association analysis also revealed meaningful co-occurrence patterns among adverse reactions. The strongest rule, “drowsiness, nausea → vomiting, dizziness,” showed a support of 0.1225, a confidence of 0.6515, and a lift of 2.823, indicating that these side effects tend to appear together more often than expected by chance. Similar high-ranking rules were concentrated around nausea, vomiting, dizziness, and headache, suggesting that adverse drug reactions often emerge as interpretable symptom clusters rather than isolated events. However, these mined associations largely reflect well-documented clinical knowledge rather than novel pharmacological discoveries. Because these rules have not yet been validated by clinical experts or pharmacists, they should be treated strictly as data-driven correlations for supplementary reference rather than actionable medical guidelines.

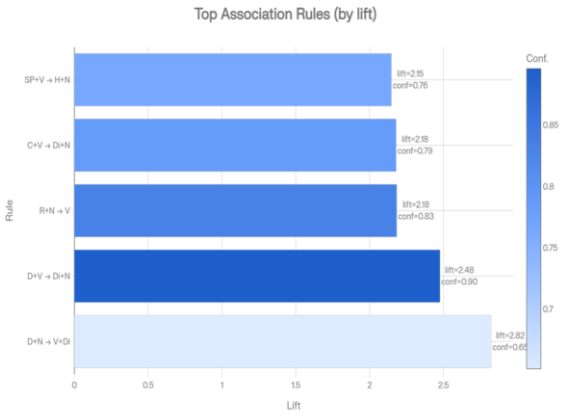


Figure 10. Top Association Rules

Finally, the Bayesian Belief Network successfully demonstrated the capacity for safe, probabilistic risk estimation. By utilizing Dirichlet smoothing parameterized with a global clinical prior, the model effectively navigated the sparse data intersections inherent in complex multi-comorbidities. When evaluated on scenarios with maximal risk factors but zero historical data representation, the system safely defaulted back to the baseline population risk (33.1%) rather than generating volatile mathematical artifacts. This validates the system's capacity to maintain rigorous safety boundaries even during edge-case query inference.

Taken together, these findings indicate that the system has practical value as a traceable medical information and pharmacovigilance support tool. The strong disease classification performance and structured knowledge base support reliable retrieval, while association rules provide useful signals for risk warning and side-effect interpretation. However, the relatively weak sentiment classification performance and the 32.9% MedDRA mapping rate for symptoms and side effects also point to future improvements, particularly through domain-specific language models and expanded medical term normalization rules.

5.4 Limitations and Future Improvement

Although the results demonstrate the usefulness of the proposed framework, several limitations remain.

First, the relatively low MedDRA mapping rate (32.9%) presents a practical limitation in our retrieval pipeline. During data integration, compound condition descriptions (e.g., 'Ulcerative Colitis, Maintenance') and user typographical errors (e.g., 'Impetig') failed to map to standardized disease_id keys in our bridge tables. Consequently, these records become 'orphaned' during strict ID-based retrieval, reducing the comprehensiveness of the chatbot's responses. Furthermore, this limits the reliability of the downstream modules, as unmapped symptom data is excluded from the association rule mining and Bayesian inference computations.

Second, the evaluation dataset was restricted to 150 manually curated queries. While intentionally designed to test high-risk edge cases and complex clinical syntax, this sample size is limited and may not fully represent the vast distribution and ambiguity of real-world patient phrasing.

Third, the current system mainly supports structured retrieval, association analysis, and lightweight probabilistic estimation, but it does not yet incorporate richer clinical validation or large-scale external benchmark testing. Additionally, despite its intended design as a public-facing health tool, the manuscript presents a system-centric evaluation and lacks a real-world usability study. Metrics such as task completion rates, user satisfaction, and whether lay users can correctly interpret the statistical outputs remain unmeasured, leaving an unvalidated gap between technical functionality and practical user experience.

Future work must therefore focus on three directions. The first is to introduce advanced biomedical NLP embeddings (e.g., BioBERT) to implement more aggressive fuzzy-matching algorithms, normalizing highly varied raw strings to improve the MedDRA mapping rate. The second is to expand the evaluation dataset to capture a wider array of colloquial symptom expressions. The third is to strengthen the evaluation framework by conducting comprehensive human-in-the-loop usability testing and securing expert pharmacological validation for the mined association rules, ensuring the chatbot is assessed not only for technical accuracy but also for practical interpretability and safety in real-world scenarios.

6 Conclusions

The retrieval-based Medical Intelligence Agent that we built has successfully bridged the gap between raw data and health awareness by integrating a structured knowledge base derived from nine heterogeneous datasets. The system manages 467 drug names, 2,655 aliases, and nearly 5,000 disease categories, enabling it to process both technical and colloquial queries. The architecture prioritizes systematic information retrieval over unrestricted generation to ensure responses remain factual and derived from verified data, functioning strictly as a knowledge-grounded framework rather than a generative model.

By implementing approximate text recognition, the agent effectively manages user typos, while specialized exclusion filters prevent the misclassification of general vocabulary. This ensures that queries regarding conditions like diabetes or medications like Metformin yield reliable, evidence-based results. Ultimately, the project met its objectives of providing fast, understandable health insights, serving as a lightweight health information support tool rather than a formal pharmacovigilance platform or a replacement for clinical professional judgement [41].

However, performance is inherently limited by the completeness of the underlying datasets, particularly the restricted clinical terminology mapping rates, as well as the current lack of clinical expert validation and real-world usability testing. Adhering to World Health Organization (2019) standards, the agent serves strictly as an educational resource and does not replace professional medical assessment or clinical diagnosis.

References

[1] Bhirud N, Athota L, Shukla VK, Pandey N, Rana A. Chatbot for healthcare system using Artificial Intelligence. Int J Adv Res Ideas Innov Technol. 2017;3(5):1727-1730.

[2] Dharwadkar R, Deshpande NA. A medical chatbot. Int J Comput Trends Technol. 2018;60:41-45. doi: 10.14445/22312803/IJCTT-V60P106.

[3] Adamopoulou E, Moussiades L. An overview of chatbot technology. In: IFIP International Conference on Artificial Intelligence Applications and Innovations; 2020. p. 373-383. doi: 10.1007/978-3-030-49186-4_31.

[4] Cheung L., Wang X, Zhang JY, Poon RKM, Lau ASM. Applications of Generative AI: A Case Study of AI Doctor. Southeast Decision Sciences Institute (SEDSI) 2025 Annual Conference.

[5] Cheung L, Cao JR, Lin Z, Sek ACH, Lau ASM. Use of a Novel Clustered Bayesian 3D-CNN Model for Triage Classification of COVID-19 Mass Casualty Incidents", Frontier in Artificial Intelligence and Applications. 2025; 412:342-258, https://doi.org/10.3233/FAIA250734.

[6] Liu CH, Pan HW, Yang LX, An YF, Hu ZW, Zhang Z, Lau, ASM. Domain-Aware Healthcare Chatbot Incorporating BERT and RAG. Frontier in Artificial Intelligence and Applications. 2025; 412:291-304. https://doi.org/10.3233/FAIA250729.

[7] Hu HY, Hu SD, Wei YH, Lau ASM. Epidemic PINNs: A chatbot based on dual data- and physical-driven epidemic prediction. Frontier in Artificial Intelligence and Applications. 2025; 412:275-290, https://doi.org/10.3233/FAIA250728.

[8] Lau ASM, Baker K, Kempf K, Grzyb K, Oke S, Tsui E, Cheung L, Slama MC, Su M. AI and Social Media Analytics for Health Systems: Understanding Consumers' Preferences in Healthcare. Adaptive Health Management Information Systems: Concepts, Cases, & Practical Applications, Chapter 13, Jones & Bartlett Publisher, September 17, 2019.

[9] Pilipiec P. Using Machine Learning for Pharmacovigilance: A Systematic Review. Pharmaceutics. 2022;14(2):266. doi: 10.3390/pharmaceutics14020266.

[10] Barreda M. Transforming healthcare with chatbots: Uses and applications—A scoping review. Digit Health. 2025. doi: 10.1177/20552076251319174.

[11] Wishart DS. DrugBank 6.0: the DrugBank Knowledgebase for 2024. Nucleic Acids Res. 2024;52:D1-D9.

[12] Kuhn M, Letunic I, Jensen LJ, Bork P. The SIDER database of drugs and side effects. Nucleic Acids Res. 2016;44(D1):D1075-D1079.

[13] International Council for Harmonisation. Medical Dictionary for Regulatory Activities: MedDRA. Available from: https://ich.org/page/meddra.

[14] Percha B. Modern clinical text mining: A review and guide. Annu Rev Biomed Data Sci. 2021;4:165-187. doi: 10.1146/annurev-biodatasci-030421-030931.

[15] Liao KP, Cai T, Savova GK, Murphy SN, Karlson EW, Ananthakrishnan AN, et al. Development of phenotype algorithms using electronic medical records and incorporating natural language processing. BMJ. 2015;350.

[16] Babu A, Boddu SB. BERT-based medical chatbot: Enhancing healthcare communication through natural language understanding. Explor Res Clin Soc Pharm. 2024;13:100419. doi: 10.1016/j.rcsop.2024.100419.

[17] Rasmy L, Xiang Y, Xie Z, Tao C, Zhi D. Med-BERT: Pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. NPJ Digit Med. 2021;4(1):86. doi: 10.1038/s41746-021-00455-y.

[18] Wu S, Roberts K, Datta S, Du J, Ji Z, Si Y, et al. Deep learning in clinical natural language processing: A methodical review. J Am Med Inform Assoc. 2020;27(3):457-470. doi: 10.1093/jamia/ocz200.

[19] Ramadhan NG, Adiwijaya, Maharani W, Gozali AA. Chronic diseases prediction using machine learning with data preprocessing handling: A critical review. IEEE Access. 2024;12:80698-80730. doi: 10.1109/ACCESS.2024.3405788.

[20] Koleck TA, Dreisbach C, Bourne PE, Bakken S. Natural language processing of symptoms documented in free-text narratives of electronic health records: a systematic review. J Am Med Inform Assoc. 2020;26(4):364-379.

[21] Agrawal R, Srikant R. Fast algorithms for mining association rules in large databases. In: VLDB '94: Proceedings of the 20th International Conference on Very Large Data Bases; 1994. p. 487-499.

[22] Venkatasubramanian V. A knowledge-graph-based pharmaceutical engineering chatbot for drug discovery. Comput Chem Eng. 2025;203. doi: 10.1016/j.compchemeng.2025.109318.

[23] Pearl J. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Elsevier Science; 1988.

[24] Horvitz E. Principles of mixed-initiative user interfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; 1999. p. 159-166.

[25] Heckerman D. A tutorial on learning with Bayesian networks. In: Learning in Graphical Models. 1998. p. 301-354.
[26] World Health Organization. WHO guideline: Recommendations on digital interventions for health system strengthening. World Health Organization; 2019 Jun 6.
[27] You Y, Gui X. Self-diagnosis through AI-enabled chatbot-based symptom checkers: User experiences and design considerations. In: AMIA Annual Symposium Proceedings; 2021. p. 1354-1363.
[28] Weizenbaum J. ELIZA—a computer program for the study of natural language communication between man and machine. Commun ACM. 1966;9(1):36-45. doi: 10.1145/365153.365168.
[29] Clark M, Bailey S. Chatbots in health care: Connecting patients to information. Can J Health Technol. 2024;1(4). doi: 10.51731/cjht.2024.846.
[30] Chow JCL, Sanders L, Li K. Impact of ChatGPT on medical chatbots as a disruptive technology. Front Artif Intell. 2023;6:1166014. doi: 10.3389/frai.2023.1166014.
[31] Fraser H, Crossland D, Bacher I, Ranney M, Madsen T, Hilliard R. Comparison of diagnostic and triage accuracy of Ada Health and WebMD symptom checkers, ChatGPT, and physicians for patients in an emergency department: Clinical data analysis study. JMIR Mhealth Uhealth. 2023;11. doi: 10.2196/49995.
[32] Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc. 2018;25(9):1248-1258.
[33] Nadarzynski T, Miles O, Cowie A, Ridge D. Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: A mixed-methods study. Digit Health. 2019;5:2055207619871808.Bickmore T, Giorgino T. Health dialog systems for patients and consumers. J Biomed Inform. 2006;39(5):556-571. doi: 10.1016/j.jbi.2005.12.004.
[34] Bickmore T, Giorgino T. Health dialog systems for patients and consumers. J Biomed Inform. 2006;39(5):556-571. doi: 10.1016/j.jbi.2005.12.004.
[35] Zhao M, Liao HT, Sun SP. An Education Literature Review on Digitization, Digitalization, Datafication, and Digital Transformation. In: 6th International Conference on Humanities & Social Science Research. Atlantis Press, Paris; 2020. p. 302-306. doi: 10.2991/assehr.k.200428.065.
[36] Zhao H, Chen H, Yang F, Liu N, Deng H, Cai H, Wang S, Yin D, Du M. Explainability for large language models: A survey. ACM Trans Intell Syst Technol. 2024;15(2):1-38.
[37] Kocielnik R, Langevin R, George JS, Akenaga S, Wang A, Jones DP, et al. Can I Talk to You about Your Social Needs? Understanding Preference for Conversational User Interface in Health. In: 3rd Conference on Conversational User Interfaces (CUI '21); 2021. doi: 10.1145/3469595.3469599.
[38] Laymouna M, Ma Y, Lessard D, Schuster T, Engler K, Lebouché B. Roles, users, benefits, and limitations of chatbots in health care: rapid review. J Med Internet Res. 2024;26:e56930.
[39] Wangsa K, Karim S, Gide E, Elkhodr M, et al. Roles, users, benefits, and limitations of chatbots in health care: rapid review. J Med Internet Res. 2024;26:e56930. doi: 10.2196/56930.
[40] Bora A, Cuayáhuitl H. Systematic analysis of retrieval-augmented generation-based LLMs for medical chatbot applications. Mach Learn Knowl Extr. 2024;6(4):2355-2374. doi: 10.3390/make6040116.
[41] Xu L, Sanders L, Li K, Chow JCL. Chatbot for health care and oncology applications using artificial intelligence and machine learning: Systematic review. JMIR Cancer. 2021;7(4):e27850. doi: 10.2196/27850.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261105

Humanizing Artificial Intelligence in Crisis: Digital Humans and Generative AI for Natural Disaster Management

Dimiter VELEV[a,1] and Plamena ZLATEVA[a,b]
[a] *University of National and World Economy, Sofia, Bulgaria*
[b] *Institute of Robotics, Bulgarian Academy of Sciences, Sofia, Bulgaria*
ORCiD ID: Dimiter Velev https://orcid.org/0000-0003-3030-1819
Plamena Zlateva https://orcid.org/0000-0002-0153-5811

Abstract. The paper explores the transformative potential of synergistic integration between Generative AI (GenAI) and Digital Humans in the context of natural disaster management. Traditional crisis notification systems often fail due to structural inelasticity and inability to overcome cognitive tunneling and panic in the threatened population. The paper proposes a new conceptual framework - "Cognitive-Affective Crisis Synergy" - in which GenAI functions as the operational computational core (the brain) and Digital Humans serve as an empathetic, multimodal interface (the face). This is a conceptual and theoretical contribution; the framework is illustrated through practical scenarios rather than empirical validation. Practical scenarios are analyzed across the three phases of crisis management: experiential learning (prevention), scalable tactical dissemination with embedded autonomous sign language translation (response), and virtual psychological triage (recovery). Special attention is paid to ethical and technological challenges, including algorithmic hallucinations, the Uncanny Valley effect under stress, legal liability assignment, and the protection of user biometric sovereignty. The paper concludes with practical recommendations for implementing RAG architectures and Edge AI systems to build a sustainable, inclusive, and humanized technological infrastructure for population protection.

Keywords. Natural Disasters, Crisis Communication, Digital Human, Generative AI, RAG Architecture

1. Introduction

The global increase of the frequency and severity of natural disaster is one of the most significant 21st century challenges. Since the 1970s their occurrence has increased more than four times, where economic losses exceed $280 billion annually [1].

Beyond the physical destruction, natural disasters trigger a secondary crisis: the collapse of communication infrastructure at the precise moment when information is most urgently needed.

The core problem is not a scarcity of data, but rather crisis communication deficit [2]: the inability of existing systems to deliver accurate, actionable information to

[1] Corresponding Author: Dimiter Velev, University of National and World Economy, 19 December 8th Str., 1700 Sofia, Bulgaria; E-mail: dgvelev@unwe.bg.

frightened individuals whose cognitive capacity has been severely impaired by severe stress. Neurological research confirms that activation of the amygdala under threat conditions reduces the prefrontal cortex's information-processing capacity by effectively narrowing the individual's attentional field - a phenomenon known as cognitive tunneling [3]. Standard broadcast messages, static web pages, and text alerts are ill-suited to penetrate this psychological barrier.

A further complication is the rapid proliferation of mis- and disinformation via social media during crisis events [4], which actively competes with official emergency communications and erodes public trust in institutional sources.

In response to these limitations, a growing body of interdisciplinary research at the intersection of Human-Computer Interaction (HCI), affective computing, and crisis management has coalesced around the concept of Humanizing AI. The term is defined here not as the superficial anthropomorphization of software, but as the design of computational systems capable of recognizing, interpreting, and responding to human emotional states and cognitive limitations in real time [5, 6].

Recent advances in two specific technologies make this vision operationally feasible: Generative AI (GenAI), particularly large language models (LLMs) and retrieval-augmented generation (RAG) architectures [7], and photorealistic Digital Humans - AI-driven 3D avatars capable of replicating nuanced human nonverbal communication through affective computing [8].

While GenAI-based crisis communication and Digital Human interfaces have each been studied independently, prior work has not addressed their combined architecture as a single sociotechnical system in high-stakes crisis contexts. Specifically, existing crisis-communication research treats message content generation and interface delivery as separate problems, without a unifying framework for how a cognitive layer (GenAI) and an affective/interface layer (Digital Human) should be coordinated across the distinct phases of disaster management, namely prevention, response, and recovery.

This paper addresses that gap by proposing the "Cognitive-Affective Crisis Synergy" framework, which explicitly models this coordination and traces its application across all three phases. GenAI functions as the operational cognitive core and Digital Humans serve as an empathetic multimodal interface. Together, they enable crisis communication systems that are both massively scalable and capable of individualized psychological support.

2. Methodological Approach

The paper adopts a conceptual synthesis methodology, integrating findings from crisis communication, affective computing, and cognitive neuroscience into a novel theoretical framework. The framework's operational implications are explored through illustrative scenario analysis across the three phases of disaster management (prevention, response, recovery). This approach is appropriate for early-stage theory-building where the object of study, the synergistic integration of two emerging technologies, has not yet been empirically instantiated in the literature, and is consistent with the conceptual/theoretical scope declared in the Abstract and revisited in the Limitations section.

The three functional components are presented in Table 1.

Table 1. Three functional components

Component	Function	Input	Output	Risk	Mitigation	Responsible Authority
GenAI Cognitive Core	Multimodal perception, synthesis, and decision support; generates contextualized instructions	Sensor/GIS telemetry, satellite/drone imagery, crowdsourced text, user geolocation & profile	Individualized, contextualized tactical instructions	Hallucination (fabricated routes/shelters)	RAG constraint restricting generation to verified database; human override protocol	Software developer / emergency headquarters
RAG Crisis Database	Single verified source of truth for all generative output	Official, authority-verified data feeds	Retrieval corpus for GenAI queries	Data pipeline compromise, cyberattack, stale data	Cryptographic signing; exclusive control by authorized emergency HQ	Data authority / emergency headquarters
Digital Human Affective Interface	Empathetic multimodal delivery (facial signals, gaze, proxemics, voice prosody)	User's vocal/facial /biometric distress signals (processed on-device)	Regulate dempathy verbal /nonverbal response calibrated to de-escalate panic	Uncanny Valley rejection; biometric privacy breach; deepfake hijacking	Non-photorealistic design or biometric-synchronized hyperrealism; on-device Edge AI only; automatic data deletion post-crisis	State operator / emergency agency (biometric protection compliance)

3. Generative AI as the Cognitive Core of Crisis Management

The theoretical framework proposed in this paper positions GenAI as the core of the crisis communication system: the layer responsible for perception, synthesis, and decision support [9]. The three primary functional capabilities that make GenAI operationally suitable for this role are analyzed: real-time multimodal data processing, dynamic generation of contextualized instructions, and elastic scaling under mass-traffic conditions.

3.1. Real-Time Multimodal Data Processing

Natural disasters generate "data tsunamis" - simultaneous, high-velocity flows of heterogeneous information that overload traditional crisis management decision-support systems [10]. They include structured telemetry (seismic sensors, Doppler weather radar, river-level gauges), unstructured visual data (high-resolution satellite imagery, real-time drone feeds), and crowdsourced textual signals (social media distress posts, SMS messages to emergency lines).

Geographic Information Systems (GIS) and crisis management platforms are primarily descriptive and they visualize incoming data but leave synthesis and interpretation to human operators. Multimodal large language models (M-LLMs), by

contrast, operate in the predictive and prescriptive domain. Through transformer self-attention mechanisms, the model continuously re-weights the relevance of each input signal against all others, enabling the identification of non-obvious cross-modal correlations. The ability to anticipate event trajectories rather than merely react to confirmed events is defined as enhanced situational awareness in emergency management [11].

3.2. Contextual Data Engineering under High Cognitive Stress

The methodological contribution of this study is the formulation of a framework for contextual data engineering under high cognitive stress, which represents a set of criteria governing how computational algorithms should filter, prioritize, and reformat information to accommodate the severely narrowed cognitive capacity of individuals in severe shock states.

The architecture operationalizes this framework through three contextualization vectors applied at the point of instruction generation:

- **Geo-spatial and temporal context:** using real-time GPS data from the user's device and the dynamic hazard perimeter maintained by the RAG database, the system calculates the user's position relative to the evolving threat and generates a specific tactical route, including real-time obstacle updates, rather than a generic directive.
- **Demographic, medical, and social profile:** using user profiles, the model adapts lexical complexity, sentence length, and instruction sequencing. For pediatric users, instructions incorporate safety gamification elements; for elderly users with limited mobility, alternative shelter-in-place protocols are prioritized.
- **Cultural-linguistic localization:** The system performs cross-linguistic transfer in real time, generating instructions in number of languages and regional dialects. In tourist-heavy areas, the model substitutes street names with salient visual landmarks, a design decision grounded in research on wayfinding under stress [12].

3.3. Elastic Scaling and Cognitive Throughput under Mass-Traffic Conditions

A critical vulnerability of human-staffed crisis communication centers is convergence behavior - the simultaneous surge in demand for information and assistance that overwhelms call center infrastructure precisely when it is most needed [13]. GenAI addresses this through elastic cloud deployment in distributed microservice containers (Kubernetes orchestration), enabling horizontal compute scaling within seconds and supporting millions of fully isolated, parallel dialogue sessions.

Each session instance applies a Summarization and Noise Reduction Filter: extracting the operational core of the user's query, filtering panic-induced verbal noise, and generating a concise, structured response calibrated to the user's current state. This architecture sustains constant cognitive throughput regardless of demand volume, providing equal access to critical, verified information across the entire affected population simultaneously.

All generative output in this architecture is governed by a Retrieval-Augmented Generation (RAG) [14]. The model is prohibited from drawing on its pre-trained

parametric knowledge and is instead forced to retrieve answers exclusively from a dynamically updated, authority-verified crisis database. The proposed RAG architecture substantially reduces, but does not eliminate, hallucination risk. Constraining generation to a verified database prevents the model from fabricating content unsupported by any source, but it does not guarantee the correctness, completeness, or currency of the retrieved content itself, nor that the correct passage is retrieved and applied to the user's specific context. Four related but distinct failure modes are defined: hallucination, retrieval error, database error, and instruction-selection error, each requiring a different mitigation strategy, shown in Table [2].

Table 2. Four failure modes.

Failure mode	Description	Origin
Hallucination	Model generates content not grounded in any retrieved source (residual risk even under RAG, e.g. from parametric leakage or overgeneralization)	Generation step
Retrieval error	Correct information exists in the database but the retriever fails to surface it (poor embedding match, query ambiguity)	Retrieval step
Database error	The underlying crisis database itself is incomplete, outdated, or contains incorrect verified-looking data	Data governance
Instruction-selection error	Multiple retrieved passages are individually correct but the model selects/combines the wrong one for the user's specific context	Generation step

4. Digital Humans as the Affective Interface of Crisis Communication

The central theoretical claim is that the modality of information delivery is not merely a design preference but a psychologically determinative variable in crisis communication, the one that directly affects compliance with life-saving instructions.

4.1. Psychophysiological Foundations: Why the Human Face Matters in Crisis

Research in evolutionary psychology and social neuroscience consistently demonstrates that humans under severe stress shift from analytical to social information processing. There is an overlap between social pain and physical threat processing in the prefrontal cortex, suggesting that social presence, the perception of another human being, is a primary regulatory mechanism for the severe stress response [15]. In the absence of such cues, the amygdala remains in hypervigilance mode, sustaining the cognitive tunneling that prevents rational compliance with instructions.

Digital Humans address this deficit through affective computing - the computational recognition, interpretation, and simulation of human emotional states. In a crisis context, the system applies a model of regulated empathy: deploying specific nonverbal signals calibrated to de-escalate distress without amplifying panic. Empirical support for the efficacy of individual nonverbal channels comes from several research traditions:

- **Gaze and mirror neuron activation:** stable, non-threatening eye contact activates mirror neuron circuits in the observer's premotor cortex, generating a felt sense of social safety [16].

- **Facial action coding (micromimicry):** the Facial Action Coding System (FACS) identifies AU1+4 (inner brow raise) as the canonical expression of empathic concern [17]. Digital Humans deploy this action unit specifically to signal deep sympathy without conveying distress.
- **Proxemic signaling:** research established that lateral head orientation signals attentiveness and non-dominance [18] - a posture the avatar adopts in response to detected vocal distress signals.

Taken together, these signals facilitate a shift from sympathetic to parasympathetic nervous system dominance, restoring the user's capacity for directed attention and instruction compliance.

4.2. Neural Voice Synthesis and Prosodic Regulation

The acoustic channel has privileged access to the human limbic system [19]. Conventional text-to-speech systems, with their flat prosody and mechanical rhythm, actively reinforce the sense of technological alienation during a disaster. Digital Humans overcome this through WaveNet-based [20] and diffusion voice synthesis architectures that enable real-time prosodic control across four parameters:

- **Fundamental frequency reduction:** A lower voice timbre is reliably associated with authority and trustworthiness across cultures [21].
- **Dynamic speech rate adaptation:** The avatar responds to rapid, fragmented user speech at a deliberate 110–120 words per minute - exploiting the phenomenon of acoustic entrainment [22] to slow the user's respiratory rate.
- **Micro-injected paralinguistic sounds:** Barely audible inhalation and exhalation cues between phrases eliminate the perception of syntheticity, a key predictor of trust rejection in human-agent interaction [23].
- **Contextual intonation shaping:** Sentence rises are suppressed during instruction delivery to project certainty; used selectively in confirmatory exchanges to signal attentiveness.

4.3. Conceptual Operational Analysis

Figure 1 presents a illustrative conceptual comparison of the operational parameters of traditional human-staffed crisis call centers and GenAI-powered Digital Human systems.

The comparison is grounded in the documented performance limitations of human crisis centers under surge conditions and the theoretical capabilities of the proposed architecture.

The data illustrate the core theoretical contribution formulated in Section 1: the union of GenAI and Digital Humans creates a new infrastructure category - an empathetic communication network with unlimited horizontal scaling. It absorbs the mass demand shock of the initial disaster phase and enables human responders to concentrate their limited capacity on complex field operations that remain beyond the competence of automated systems.

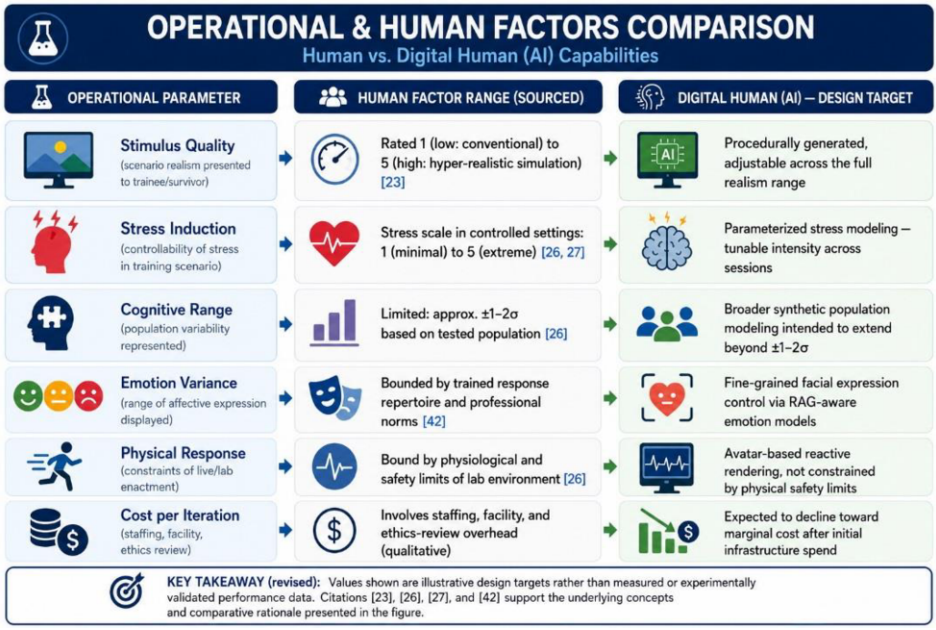


Figure 1. Operational parameters: Human Crisis Centers vs. GenAI-powered Digital Humans

5. Synergy in Action: Scenarios Across the Disaster Life Cycle

The theoretical framework in operational reality is determined by analyzing the proposed system's application across the three phases of the disaster management life cycle [24]: prevention, response, and recovery. Each phase is analyzed through documented failure modes in existing systems, illustrated where possible with reference to real-world disaster events. Figure 2 provides a structural overview.

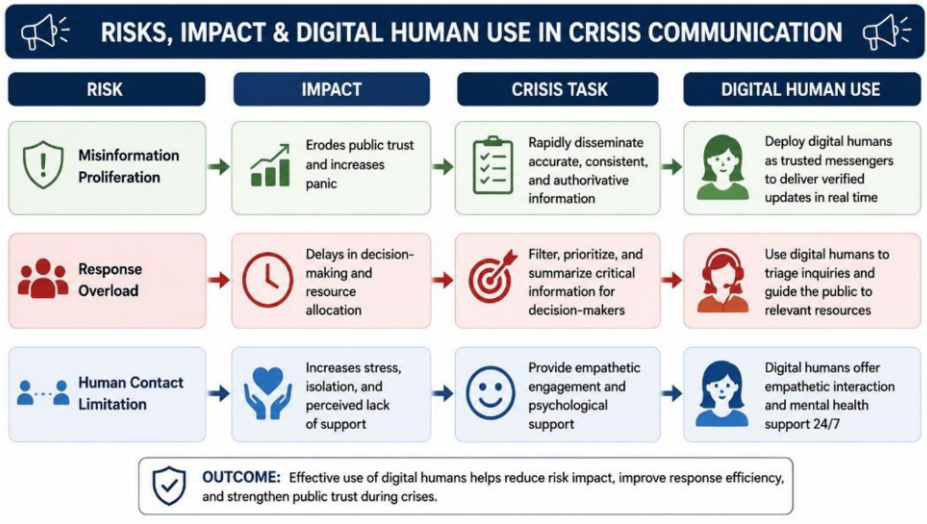


Figure 2. System role distribution across the three-phase disaster lifecycle.

5.1. Prevention Phase: From Passive Instruction to Experiential Crisis Learning

A persistent structural failure of conventional disaster preparedness programs is low knowledge retention under stress conditions, a phenomenon known as crisis amnesia: the inability to recall procedural safety information precisely when it is needed most [25]. Meta-analytic evidence indicates that passive instructional formats (brochures, public service announcements) produce retention rates of approximately 10–15% under stress conditions, compared to 65–75% for simulation-based experiential learning [26, 27].

The proposed system addresses this deficit through what is termed Experiential Crisis Learning (ECL) - a closed-loop simulation ecosystem in which GenAI generates a hyper-local, personalized disaster scenario based on the user's verified residential and structural data. The Digital Human instructor, designed as a civil protection officer, guides the user through the simulation interactively, evaluating verbal responses in real time and providing immediate behavioral correction through expressive avatar feedback.

A relevant real-world precedent is Japan's ShakeOut earthquake drill program, which has demonstrated measurable reductions in injury rates among trained populations [28]. The ECL model extends this approach by personalizing scenarios to individual household configurations and delivering them continuously via mobile application, rather than through infrequent mass drills.

5.2. Response Phase: RAG-Filtered Dissemination and Inclusive Broadcasting

The Kahramanmaraş earthquake in Turkey in 2023, which killed over 50,000 people across ten provinces, provides a critical case study in the failure of response-phase crisis communication. Post-event analysis by the UN Office for the Coordination of Humanitarian Affairs [29] identified three systemic failures: collapse of telecommunications infrastructure within hours of the initial shock; rapid spread of false evacuation routes through social media; complete exclusion of an estimated 1.2 million foreign nationals from Turkish-language official communications.

The proposed system directly addresses each of these failure modes. The RAG core provides a single, authority-verified information channel that immediately retracts or revises any instruction affected by new operational data - road closures, updated hazard perimeters, structural collapses. Population segmentation by real-time geolocation enables micro-targeted instruction delivery, with content automatically adapted to the recipient's language profile and local topographic familiarity.

For populations with disabilities, the system realizes the Inclusive Design principles articulated in UN CRPD General Comment No. 2 [30]: Digital Humans broadcast emergency messages with simultaneous neural sign language translation with facial expression maintained at the precision level required to disambiguate sign meaning, and provide detailed audio wayfinding for visually impaired users. Edge AI architecture ensures continuity of service under telecommunications failure: lightweight model versions pre-loaded on user devices render a functional avatar via satellite text relay, a capability demonstrated in analogous form during the 2021 Hurricane Ida response in Louisiana 2022 [31].

Edge AI needs operational assumptions, such as: Device capability: minimum GPU/NPU power, which is required to render a lightweight avatar and run on-device inference; Battery constraints: continuous on-device inference and rendering is assumed with duration before thermal/battery throttling, and a low-power fallback mode; Model update mechanism is required in which the pre-loaded lightweight model and crisis

database are kept current when connectivity is intermittent;Offline language support: on-device models must ship with pre-loaded language packs for the deployment region; Failure modes: measures must be fined when the device lacks sufficient capability, battery is exhausted, or the local model is out of date.

5.3. Recovery Phase: Psychological Triage and Bureaucratic Navigation

Post-disaster psychological morbidity is extensively documented. Following the 2011 Great East Japan Earthquake, researchers reviewing comparable large-scale disaster data found PTSD prevalence rates of 30–40% among directly affected populations, with depression rates of 20–30% [32, 33]. Critically, the ratio of affected individuals to available mental health professionals in disaster contexts typically exceeds 500:1, rendering traditional therapeutic coverage structurally impossible in the recovery phase.

The proposed system's recovery-phase role is explicitly bounded to psychological first aid and distress screening, not clinical triage or diagnosis. The Digital Human, fine-tuned on validated psychological first aid protocols [34], provides grounding techniques, cognitive-behavioral exercises, and breathing practices in a confidential, on-demand setting. Clinical functions including suicide-risk assessment, diagnostic triage, and any escalation decision remain the responsibility of licensed human clinicians. The system's role is limited to continuous, non-diagnostic monitoring of vocal and facial distress indicators as a screening trigger: upon detection of markers associated with suicidal ideation or psychotic features, the session is escalated immediately to an on-duty human clinician who receives a structured summary and makes the clinical determination. Deployment of this screening function requires prior clinical validation of the detection model's sensitivity and specificity, human-in-the-loop oversight for every escalation, and a governance protocol defining response-time guarantees and liability for missed or false escalations.

The second recovery-phase function addresses what survivors consistently report as a source of secondary trauma: the bureaucratic complexity of insurance claims, housing applications, and compensation documentation in the immediate aftermath of property loss [35]. The system deploys computer vision to parse official documents, explains their content in accessible language, and auto-completes required forms from the user's verbal responses - routing completed documents directly to the relevant institutions.

6. Ethical Considerations and Challenges

Autonomous sociotechnical systems deployed in life-safety contexts introduce ethical, legal, and ontological challenges that are qualitatively distinct from those encountered in commercial AI applications. Drawing on the emerging regulatory framework of the EU AI Act [36, 37] and relevant research traditions, four principal challenge domains are analyzed below. Figure 3 provides a consolidated risk register with severity assessments and mitigation strategies.

6.1. Epistemic Reliability: Hallucination Risk and Liability Assignment

The defining epistemic vulnerability of large language models in high-stakes contexts is the phenomenon of hallucination - the generation of factually incorrect content with high linguistic confidence, arising from the statistical nature of next-token prediction [38].

In crisis communication, a hallucinated evacuation route or a fabricated shelter location is not a recoverable error: it is a potential cause of mass casualties. This risk is categorically incompatible with deployment in life-safety contexts unless architectural safeguards eliminate it.

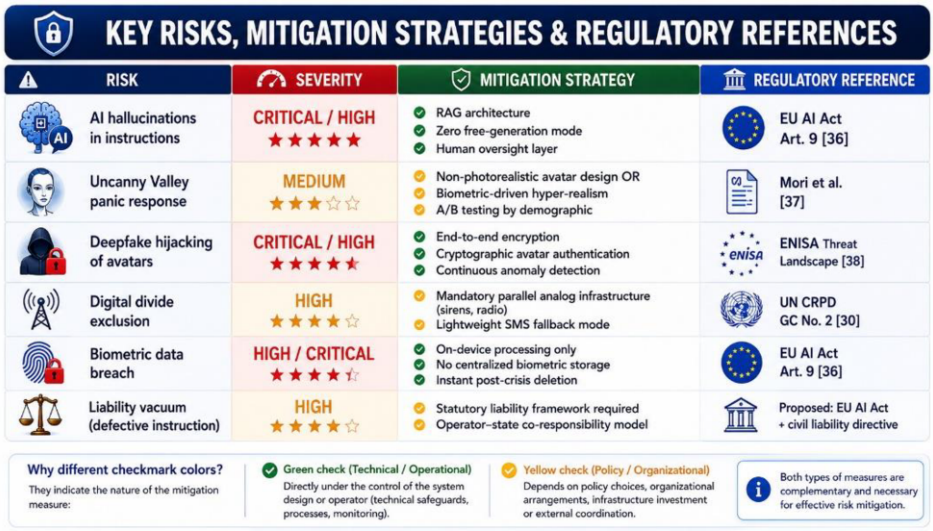


Figure 3. Ethical and technological risk register for crisis-deployed AI systems.

The proposed RAG architecture addresses this by enforcing a strict epistemic boundary: the generative core is prohibited from accessing its pre-trained parametric knowledge and is forced to retrieve all outputs from a dynamically updated, cryptographically signed crisis database under the exclusive control of authorized emergency headquarters.

The RAG constraint does not resolve the parallel legal question of liability assignment, which represents a gap explicitly identified in the original text of this paper. When a system operating under RAG constraints delivers a defective instruction, whether through a database error, an inference failure, or a cyberattack on the data pipeline, existing law provides no clear allocation of responsibility between the software developer, the state operator, and the data authority. The EU AI Act imposes conformity assessment and risk management obligations on high-risk AI operators but explicitly defers civil liability to Member State law. The proposed EU AI Liability Directive [39] introduces a rebuttable presumption of causality for AI-induced harm but requires transposition that has not yet occurred in most jurisdictions. The study recommends a framework between developers and government operators, with mandatory indemnification provisions and clear human override protocols.

6.2. The Uncanny Valley Effect in Stress Conditions

Mori's Uncanny Valley hypothesis [40] positions that as a robotic or virtual entity approaches but does not achieve human perceptual realism, observer affect drops sharply into a region of discomfort and revulsion. This effect is substantially amplified under stress conditions: amygdala hyperactivation increases sensitivity to threat-relevant perceptual anomalies [41]. This means that subtle avatar defects, such as lip-sync latency

above approximately 80ms, unnatural corneal reflections, or rigid cervical kinematics, may be processed as deception signals, triggering secondary panic or complete instruction rejection.

The study proposes two design strategies to navigate this risk, which should be selected based on deployment context and user demographic:

- **Non-photorealistic avatar design:** Deliberate stylization below the Uncanny Valley threshold. The brain immediately categorizes the entity as artificial intelligence, suppressing the search for micro-defects while preserving the affective value of voice prosody and expressive facial behavior. This approach is supported by character animation research, which shows sustained user engagement with stylized non-humanoid agents [42, 43].
- **Biometric-driven hyper-realism:** Neural synchronization of the avatar's micro-saccades, pupil dilation, and respiratory rhythm with real-time psychophysiological signals from the user. By mirroring the user's own physiological state, the avatar functions as a stable cognitive anchor rather than an alienating simulacrum. This approach requires robust real-time biometric processing.

6.3. Digital Divide, Biometric Sovereignty, and Cybersecurity

The deployment of computationally intensive GenAI systems in crisis contexts risks reproducing and amplifying existing patterns of social exclusion. The ITU Global Connectivity Report [44] documents that approximately 2.7 billion people remain unconnected to the internet, with the highest concentrations in rural and low-income populations, which are precisely those with the highest disaster vulnerability index. High-fidelity Digital Human rendering requires device GPU capacity and network bandwidth unavailable to these populations. If government investment in AI crisis communication displaces funding for analog universal coverage systems, such as sirens, broadcast radio, physical rescue patrols, then the outcome is a net reduction in protection for the most vulnerable. The study endorses the UN Sendai Framework's principle [45] that technological innovation in disaster risk reduction must complement, not replace, universal coverage mechanisms.

The system's empathic adaptation capabilities require continuous access to sensitive biometric data: real-time facial expression analysis, vocal biomarker extraction, and precise geolocation. Biometric data in high-risk AI systems is subject to heightened protection requirements. The study proposes a strict Privacy-by-Design architecture, where all biometric processing must occur exclusively on the user's device via Edge AI inference [46]. No biometric data may be transmitted to or stored on centralized servers and all locally processed data must be deleted automatically upon conclusion of the crisis event, without exception. A successful breach of a centralized avatar control system could enable mass Deepfake Hijacking: the broadcast of false evacuation instructions to redirect the population toward danger zones, which represents a scenario with direct national security implications. Distributed on-device architecture eliminates this attack surface entirely.

7. Conclusion

7.1. Research Summary and Theoretical Contributions

This paper has presented and defended the thesis that the synergistic integration of GenAI and Digital Humans constitutes a disruption in crisis communication- the one that is termed the Cognitive-Affective Crisis Synergy framework. The framework addresses the three structural failure modes of existing crisis communication systems identified in the opening section: the cognitive overload of threatened individuals (addressed by GenAI's contextual data engineering); the uniformity and impersonality of broadcast communications (addressed by Digital Humans' affective interface); and the absence of scalable empathic presence (addressed by the synergistic architecture as a whole).

Four scientific contributions were advanced and substantiated across the paper:

- **Theoretical:** The Cognitive-Affective Crisis Synergy paradigm extends] Crisis Communication Theory by formalizing the principle of automated empathy at scale and defining the operational roles of GenAI (cognitive core) and Digital Humans (affective interface) within a unified system architecture.
- **Methodological:** A framework for contextual data engineering under high cognitive stress, specifying three contextualization vectors (geo-spatial, demographic, and cultural-linguistic) that govern the generation of personalized crisis instructions adapted to the narrowed cognitive capacity of individuals in severe shock states.
- **HCI:** Theoretical guidelines for navigating the Uncanny Valley effect in critical conditions, proposing two design strategies, non-photorealistic stylization and biometric hyper realism, with selection criteria based on deployment context.
- **Social-inclusive:** A framework for automated accessibility through multimodal real-time translation, which satisfies the system inclusive design requirements in the obligations of UN CRPD General Comment No. 2 and the UN Sendai Framework 2015.

7.2. Practical Recommendations for Implementation

Three recommendations are derived for policymakers, government emergency management agencies, and technology developers, each addressing a critical implementation risk identified in Section 5:

- **Mandatory RAG-only architecture (addressing hallucination risk):** Regulatory frameworks, specifically EU AI Act Annex III, should prohibit free-generation mode in all crisis-deployed AI systems and require that all outputs be retrieved from cryptographically signed, dynamically updated, authority-verified databases. Human oversight protocols must define escalation thresholds for system override.
- **Inclusive and psychologically calibrated design (addressing Uncanny Valley and digital divide risks):** Avatars must be validated against Uncanny Valley thresholds for each target before their deployment. All systems must provide real-time sign language translation, which is to be compliant with national linguistic standards, multilingual voice support, and a lightweight SMS mode of operation for low-connectivity users.

- **Edge AI as universal baseline infrastructure (addressing telecommunications failure and digital divide):** Lightweight model versions and avatar graphical assets must be pre-integrated at the operating system level on user devices, enabling autonomous offline operation via satellite emergency text relay. This capability should be treated as a public infrastructure standard rather than an optional feature, analogous to the mandatory installation of emergency broadcast receivers in vehicles.

7.3. Limitations of the Present Study

Intellectual honesty requires explicit acknowledgment of the present study's limitations, which circumscribe the scope of its claims:

- **Empirical gap:** The analysis is entirely theoretical and conceptual. No controlled trials, deployment data, or user studies are reported. The practical scenarios in Section 4 describe hypothetical applications of the proposed architecture.
- **Technology readiness:** The affective computing capabilities attributed to Digital Humans, real-time emotional recognition at clinical precision, biometric-synchronized prosody, and seamless offline rendering, represent the current research frontier but are not yet deployed at production scale in crisis contexts. Implementation timelines and cost structures remain uncertain.
- **Cross-cultural validity:** The trust calibration assumptions underlying the avatar design recommendations are derived primarily from Western HCI research traditions. Significant cross-cultural variation in the perception of digital human realism, authority, and empathy has been documented and is not systematically addressed in the present framework.
- **Scope of ethical analysis:** The ethical framework in Section 5 focuses on technical and legal risks. Broader societal questions, such as the long-term effects of AI-mediated crisis communication on institutional trust, the political economy of crisis AI procurement, and the potential for state surveillance via crisis biometric collection, should have dedicated treatment beyond the scope of this paper.

7.4. Future Research Directions

The limitations identified above define a structured agenda. Figure 4 maps the priority research directions to proposed methodologies and expected contributions.

The agenda will require interdisciplinary cooperation among IT professionals, cognitive psychologists, emergency management experts, legal researchers, and the communities affected by crisis events.

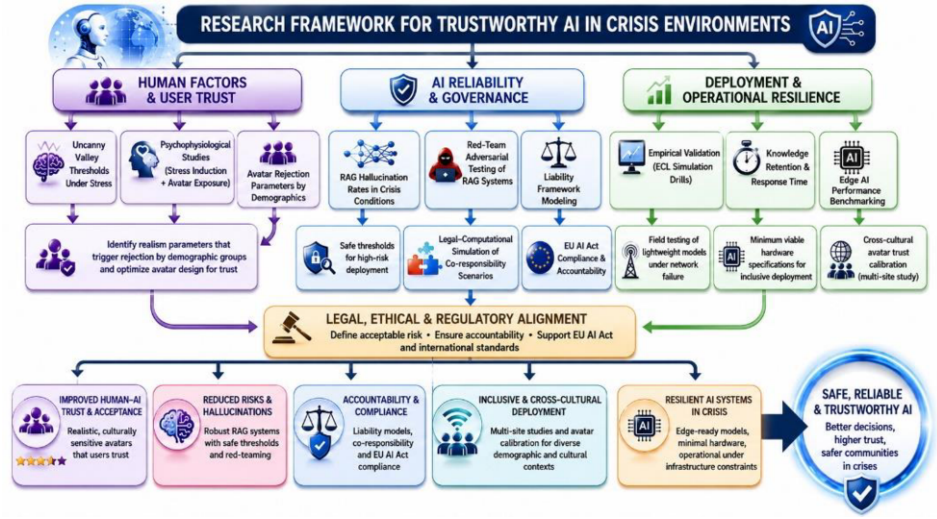


Figure 4. Future research agenda: priority directions, methodologies, and expected contributions

Acknowledgments

This work has been carried out in the framework of the Project "Smart Risk Management for Business from Adverse Events and Natural Disasters", financially supported by the National Science Fund of Bulgaria, Grand № KP-06-H55/5 from 15.11.2021.

References

[1] Centre for Research on the Epidemiology of Disasters (CRED). EM-DAT: The International Disaster Database. Université catholique de Louvain. 2023, https://www.emdat.be
[2] Information Resources Management Association. Emergency and Disaster Management: Concepts, methodologies, tools, and applications. IGI Global, 2019, doi: 10.4018/978-1-5225-6195-8
[3] Arnsten A. Stress signalling pathways that impair prefrontal cortex structure and function. Nat Rev Neurosci. 2009;10:410–422, doi: 10.1038/nrn2648
[4] Starbird K, Maddock J, Mania O, Achterman P, Mason RM. Rumors, False Flags, and Digital Vigilantes: Misinformation on Twitter after the 2013 Boston Marathon Bombing. Proceedings of the iConference, 2014, University of Illinois Urbana-Champaign, doi: 10.9776/14308
[5] Picard RW. Affective Computing. MIT Press, 1997, doi: 10.7551/mitpress/1140.001.0001
[6] Hinss MF, Brock AM, Roy RN. Cognitive effects of prolonged continuous human-machine interaction: The case for mental state-based adaptive interfaces. Front Neuroergon. 2022 Aug;3:935092, doi: 10.3389/fnrgo.2022.935092
[7] Lewis P, Perez E, Piktus A, et al. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. Advances in Neural Information Processing Systems. 2020;33:9459–9474, doi: 10.48550/arXiv.2005.11401
[8] Calvo RA, D'Mello SK, Gratch J, Kappas A, editors. The Oxford handbook of affective computing. Oxford University Press, 2015, doi: 10.1093/oxfordhb/9780199942237.001.0001
[9] Sellnow TL, Seeger MW. Theorizing Crisis Communication, 2nd edition, Wiley-Blackwell, 2020.
[10] Chen R, Sharman R, Rao HR, Upadhyaya SJ. Coordination in Emergency Response Management. Communications of the ACM. 2014;51(5):66–73, https://cacm.acm.org/research/coordination-in-emergency-response-management/
[11] Vaswani A, Shazeer N, Parmar N, et al. Attention Is All You Need. In: Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS), Long Beach, CA, USA, 2017, doi: 10.48550/arXiv.1706.03762

[12] Golledge RG. Wayfinding Behavior: Cognitive Mapping and Other Spatial Processes. Johns Hopkins University Press, 1999, https://www.press.jhu.edu/books/title/3013/wayfinding-behavior
[13] Heide EA. The Importance of Evidence-Based Disaster Planning. Annals of Emergency Medicine. 2006;47(1):34-49, doi: 10.1016/j.annemergmed.2005.05.009
[14] James A, Trovati M, Bolton S. Retrieval-Augmented Generation to Generate Knowledge Assets and Creation of Action Drivers. Applied Sciences. 2025;15(11):6247, doi: 10.3390/app15116247
[15] Eisenberger NI, Lieberman MD. Why rejection hurts: a common neural alarm system for physical and social pain. Trends in Cognitive Sciences. 2004;8(7):294-300, doi: 10.1016/j.tics.2004.05.010
[16] Prinsen J, Alaerts K. Eye contact enhances interpersonal motor resonance: comparing video stimuli to a live two-person action context. Social Cognitive and Affective Neuroscience. 2019 Sept;14(9):967-976, doi: 10.1093/scan/nsz064
[17] Ekman P, Friesen WV, Hager JC. Facial Action Coding System: Manual and Investigator's Guide. Salt Lake City, UT: A Human Face, 2002, https://www.paulekman.com/facial-action-coding-system/
[18] Hall ET. The Hidden Dimension (Anchor Books a Doubleday Anchor Book). Anchor, 1990, https://www.amazon.com/Hidden-Dimension-Anchor-Books-Doubleday/dp/0385084765
[19] LaBar KS. Fear, learning, and the amygdala: a personal reflection in honor of Joseph LeDoux. Cerebral Cortex. 2025;35(1):5-10, doi:10.1093/cercor/bhae394
[20] Oord A, Dieleman S, Zen H, et al. WaveNet: A Generative Model for Raw Audio. arXiv:1609.03499, 2016, doi: 10.48550/arXiv.1609.03499
[21] Shiramizu VKM, Lee AJ, Altenburg D, et al. The role of valence, dominance, and pitch in perceptions of artificial intelligence (AI) conversational agents' voices. Sci. Rep. 2022;12:22479, doi: 10.1038/s41598-022-27124-8
[22] Berger J, Turow G. (editors) Introduction to entrainment and cognitive ethnomusicology. In: Music, Science and the Rhythmic Brain. Routledge, 2011, https://www.routledge.com/Music-Science-and-the-Rhythmic-Brain-Cultural-and-Clinical-Implications/Berger-Turow/p/book/9780415709484
[23] Miller EJ, Foo YZ, Mewton P, Dawel A. How do people respond to computer-generated versus human faces? A systematic review and meta-analyses. Computers in Human Behavior Reports. 2023;10:100283, doi: 10.1016/j.chbr.2023.100283
[24] Coppola DP. Introduction to International Disaster Management. Butterworth-Heinemann, 4th edition, 2020, https://www.sciencedirect.com/book/monograph/9780128173688/introduction-to-international-disaster-management
[25] Cohn MA, Fredrickson BL, Brown SL, Mikels JA, Conway AM. Happiness unpacked: Positive emotions increase life satisfaction by building resilience. Emotion. 2009;9(3):361-368, doi: 10.1037/a0015952
[26] Starcke K, Brand M. Decision making under stress: A selective review. Neuroscience & Biobehavioral Reviews. 2012;36(4):1228-1248, doi: 10.1016/j.neubiorev.2012.02.003
[27] Sarmiento LF, Lopes da Cunha P, Tabares S, Tafet G, Gouveia A Jr. Decision-making under stress: A psychological and neurobiological integrative model. Brain Behav Immun Health. 2024 Apr;38:100766, doi: 10.1016/j.bbih.2024.100766
[28] Freddi F, Galasso C, et al. Innovations in earthquake risk reduction for resilience: Recent advances and challenges. International Journal of Disaster Risk Reduction. 2021;60:102267, doi: 10.1016/j.ijdrr.2021.102267
[29] OCHA. Turkey-Syria Earthquake: Situation Report No. 12. UN Office for the Coordination of Humanitarian Affairs, 2023, https://reliefweb.int/report/turkiye/turkiye-2023-earthquakes-situation-report-no-12-30-march-2023-entr
[30] UN Committee on the Rights of Persons with Disabilities. General Comment No. 2: Article 9 (Accessibility). CRPD/C/GC/2, 2014, https://docs.un.org/en/CRPD/C/GC/2
[31] FEMA. Hurricane Ida After-Action Report. Federal Emergency Management Agency, U.S. Department of Homeland Security, 2022, https://www.nrc.gov/docs/ML2209/ML22097A138.pdf
[32] Rostami ZC, Daliri S, Rohani-Rasaf M, Karimi A, Zare F. Post-Traumatic Stress Disorder after Disaster and Mass-Casualty Incidents in Developed and Developing Countries: A Meta-Analysis Study. Iran J Psychiatry. 2025 Jul;20(3):383-404, doi: 10.18502/ijps.v20i3.19044
[33] Sonmez D, Hocaoglu C. Post-Traumatic Stress Disorder after Natural Disasters: A Review. Duzce Med J. 2023;25(2):103-114, doi:10.18678/dtfd.1277673
[34] WHO. Psychological First Aid: Guide for Field Workers. World Health Organization, 2011, https://www.who.int/publications/i/item/9789241548205
[35] Galea S, Brewin CR., Gruber M, et al. Exposure to hurricane-related stressors and mental illness after Hurricane Katrina. Archives of General Psychiatry. 2005;64(12):1427–1434, doi: 10.1001/archpsyc.64.12.1427
[36] European Parliament. Regulation (EU) 2024/1689 of the European Parliament and of the Council - Artificial Intelligence Act. Official Journal of the European Union, 2024, https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

[37] ENISA. Threat Landscape 2023. EU Agency for Cybersecurity, 2023, https://www.enisa.europa.eu/sites/default/files/publications/ENISA%20Threat%20Landscape%202023.pdf
[38] Maynez J, Narayan S, Bohnet B, McDonald R. On Faithfulness and Factuality in Abstractive Summarization. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL-2020), July 2020, pp. 1906–1919, Online, Association for Computational Linguistics, doi: 10.18653/v1/2020.acl-main.173
[39] European Commission. Proposal for a Directive on Adapting Non-Contractual Civil Liability Rules to Artificial Intelligence. COM (2022) 496, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52022PC0496
[40] Mori M, MacDorman KF, Kageki N. The Uncanny Valley [From the Field]. In: IEEE Robotics & Automation Magazine, vol. 19, no. 2, pp. 98-100, June 2012, doi: 10.1109/MRA.2012.2192811
[41] Meijer S, Carpino E, Kop BR, Lam J, de Voogd LD, Roelofs K, Verhagen L. The human amygdala in threat learning and extinction. Science Advances. 2026;12(13), doi: 10.1126/sciadv.aea8233
[42] Natale AFdi, Simonetti ME, Rocca SL, Emanuela Bricolo E. Uncanny valley effect: A qualitative synthesis of empirical research to assess the suitability of using virtual faces in psychological research. Computers in Human Behavior Reports. 2023;10:100288, doi: 10.1016/j.chbr.2023.100288
[43] Hepperle D, Purps CF, Deuchler J, et al. Aspects of visual avatar appearance: self-representation, display type, and uncanny valley. Vis Comput. 2022;38:1227-1244, doi: 10.1007/s00371-021-02151-0
[44] ITU. Global Connectivity Report - 2022. International Telecommunication Union, 2022, https://www.itu.int/itu-d/reports/statistics/global-connectivity-report-2022/
[45] UNDRR. Sendai Framework for Disaster Risk Reduction 2015–2030. United Nations Office for Disaster Risk Reduction, 2015, https://www.undrr.org/publication/sendai-framework-disaster-risk-reduction-2015-2030
[46] Kisselburgh L, Beever J. The Ethics of Privacy in Research and Design: Principles, Practices, and Potential. In: Knijnenburg BP, Page X, Wisniewski P, Lipford HR, Proferes N, Romano J. (eds), Modern Socio-Technical Perspectives on Privacy. Springer Cham, 2022, doi: 10.1007/978-3-030-82786-1_17

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261106

Integrating Lexicon-based Approach and Machine Learning Method for Content Analysis: A Conceptual Framework

Sitong LI[a,2] and Adela S.M. LAU[b,c,1]
[a] *Shanghai Shenghua Zizhu Academy, China*
[b] *Data Science Lab, Department of Statistics and Actuarial Science, School of Computing and Data Science, The University of Hong Kong, China*
[c] *Dr Adela Education Limited*
ORCiD ID:
Sitong Li https://orcid.org/0009-0008-8380-7075
Adela S.M. Lau https://orcid.org/0000-0001-5918-8309

Abstract. Traditional content analysis uses natural language processing, linguistics and text analysis methods for analyzing the text. The challenges of content analysis are the understanding of word meanings, synonyms, colloquialism, slang words, and domain-specific language. Although lexicon method can overcome these problems by using a dictionary, it cannot understand the irony words, slang, new terminology, or implicit meanings of the text that make the classification not accuracy. Although machine learning methods overcome the limitation of lexicon method, it cannot understand the context of the text for analysis. Transformer and self-attention overcome this limitation; it did not address the semantic meaning of the lexicon words to classify the text with it semantic meaning in the output. Therefore, novel approach of lexicon-based machine learning context analysis method was proposed. The conceptual framework of the novel analysis method was proposed. An example of applying it to mathematics for the conceptual design of the content analysis was done. The pros and cons of various machine learning methods for classification were discussed.

Keywords. user-defined lexicon, content-based analysis, artificial intelligence.

1. Introduction

Computational linguistics and text analysis use machine learning and AI methods to extract, relate, and conceptualize words and phrases from the text for understanding the meaning of the text [1][2]. It uses natural language processing (NLP) to extract keywords and phrases from the text, and to analyze the syntactic, lexical, and semantic meanings of it. In order to understand the word meaning and the word relationships, machine learning and lexicon-based approach are commonly used. However, there are some

[1] Corresponding Author: Adela S.M. Lau, Data Science Lab, Department of Statistics and Actuarial Science, Run Run Shaw Building, The University of Hong Kong, Pokfulam Road, Hong Kong; Email: adelalau@hku.hk

limitations of these methods. This paper aims to discuss the strengths and limitations of the machine learning and lexicon-based methods for content analysis, and to propose a conceptual framework for implementing the novel lexicon-based machine learning model for text classification with its semantic meaning in the output. In the following chapter, the review of the traditional content analysis methods will be discussed. A novel lexicon-based machine learning text classification model will then be discussed. An example, for the conceptual design of classifying Mathematics topic will be used to demonstrate how the novel approach can be used. The pros and cons of various machine learning methods for topic classification will be finally discussed.

2. Literature Review

Traditional content analysis can be classified into machine learning approach [2][3] and lexicon-based approach [4][5]. A lexicon-based approach matches the words and phrases with a pre-defined dictionary to understand the meaning of the text. It requires no training data to train the model that makes the deployment more efficient [6]. It uses a human defined dictionary for defining a set of words for classification [1][2] that makes it with a higher accuracy of classification and easy to manage words in dictionary [6][7]. It can be easier to customize the dictionary by adding domain-specific words to the existing one to make a domain specific text classifier [6]. However, it cannot understand the irony words that may make errors in text classification. It requires constant manual updates to maintain the accuracy. It cannot understand the content with the slang, new terminology, or implicit meanings of the text that make the classification not accuracy.

Machine learning learns the pattern of words, and has supervised and unsupervised learning methods for classification. The supervised learning methods include decision tree, linear classifier, rule-based approach, and probabilistic approach for classification [3][4][8]. The unsupervised learning methods include clustering and dimensional reduction. Machine learning extracts the feature words using Parts of Speech (POS). It uses monogram, bi-gram, or n-gram to analyze the semantic meaning of the text. A bag-of-words and tf-idf are the commonly used method for text classification. Some other machine learning methods, such as Navie Bayes and Vector and Maximum Entropy, are also commonly used in text classification. The strengths of machine learning are its capability to analyze a huge volume of text in a short time with the trained model. It can train the model with different languages, slang, and domain-specific words. It can retrain with new data with little effort. It can understand the fatigue and subjective words by adding rules, and detect word patterns for classification. Therefore, machine learning methods overcome the limitations of the lexicon-based approach for text classification. However, the limitation of the machine learning method requires huge and human intensive labelled datasets to train the model that makes a high training time and cost. The same text in different context cannot be differentiated by machine learning that results in misclassification.

Although some researchers have used lexicon-based approach and machine learning approach, such as k-Nearest Neighborhood [9], Conditional Random Field [10], Hidden Markov Model [11], Multi-Dimensional Classification [12], or Chain Optimization [13], together for text classification, it cannot classify the text effectively with the domain specific problem. The researchers only focus on extracting the lexicon words in the content and use the document content to train the machine learning model for text analysis or classification. It only learns the lexicon words together with the words in the

document without understanding the context of the document. A context-based content analysis is required. Transformer and self-attention mechanism [14] is the recent context-based analysis method. Researchers used the attention mechanism to weight the context words and lexicon words to train the model for improving the accuracy of the text classification for better to solve the domains specific problem. However, it is time consuming and requires extensive document to train a better model. It did not address the semantic meaning of the lexicon words itself to train the model to classify the text with the semantic exploring of the classified text. Therefore, a lexicon-based machine learning text classification method is proposed for the context-based content analysis for semantic text classification. It uses lexicon-based approach to define dictionary by human for creating knowledge domain context, and uses a multi-level approach to design the machine learning models with the lexicon words and the neighbor words of the lexicon words to analyze the text in the document context for text classification with semantic exploring.

3. A Conceptual framework of the Novel Lexicon-based Machine Learning Text Classification Method

Traditional lexicon-based text classifcation only consider the lexicon structure of a specific domain for analysis. It did not address the lexicon text itself in different context that causes misclassification of the text. Therefore, in this study, a set of domain specific word categories were used to break down the context of a document in a hierarchical structure. The lexicon words of different specific domains are trained with a multi-level machine learning model to classify lexicon words into different topics with its semantic exploring to improve the text classification.

For example, in mathematics, it has different topics (e.g. rational numbers) and sub-topics (e.g. positive and negative numbers, number line, opposite numbers, absolute value, arithmetic operations with rational numbers, exponentiation). The lexicon words of fraction, integers, ratio, numerator, denominator, and decimal are used to classify the topic of rational number. The context of the lexicon word, such as fraction, can be a theory, a concept or an example for the rational number topic (i.e. the context of the lexicon words). To further understand the lexicon word's semantic meaning and application to the topic, a multiple-level machine learning model by using the neighbor words of the lexicon words can be used to classify the lexicon words to be a theory, concepts or example of the topic.

Therefore, an overview of the lexicon-based machine learning text classification method is shown in Figure 1. The steps of the model design and implementation is presented in Figure 2. The first step is that a user-defined dictionary is inputted by users to define the specific words of a domain knowledge (e.g. rational number). The raw data of the domain document is collected. Next, the data is pre-processed by removing the stop word, special characters, symbols, colloquialisms, and slang words. Fourth, the multi-level machine learning model is designed and trained with the domain-specific words and their relationships. Finally, the model is trained and tested.

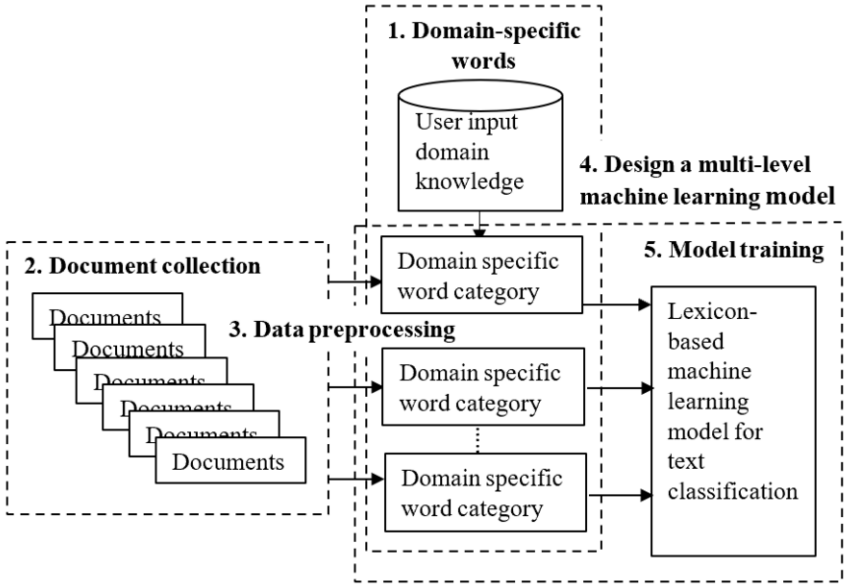


Figure 1. An overview of the methodologies of lexicon-based machine learning text classification method.

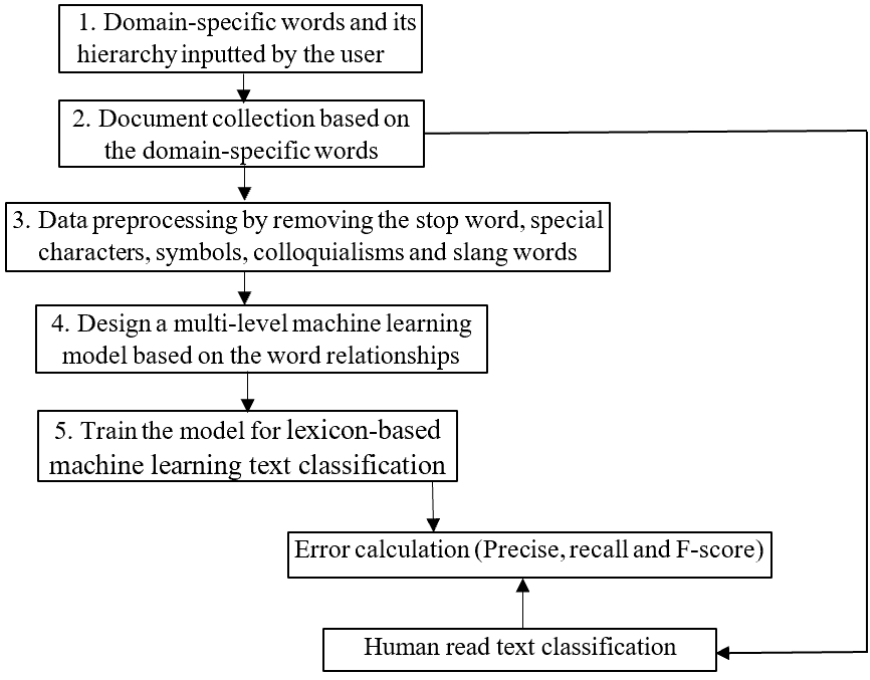


Figure 2. Steps for the lexicon-based machine learning model design.

4. Example on the lexicon-based machine learning method for topic classification

In order to classify the domain knowledge into different words, the mathematics topic was used as an example. User defined the topics and sub-topics as shown in Table 1.

Table 1. The mathematical topics and subtopics of the 7th class.

Year	Term	Topic	Second Topic/Syllabus
Year 7	1st term	Rational numbers	Positive and negative numbers, number line, opposite numbers, absolute value, arithmetic operations with rational numbers, exponentiation.
Year 7	1st term	Addition and subtraction of integral expressions	Algebraic expressions, monomials, polynomials, and combining like terms
Year 7	1st term	Linear equation in one variable	Definition of an equation, properties of equality, solving linear equations in one variable, and real-world application problems.
Year 7	1st term	Introduction to Geometric Figures	Three-dimensional and two-dimensional figures; points, lines, planes, and solids; angles and their measurement.
Year 7	2nd term	Intersecting lines and parallel lines	Adjacent supplementary angles, vertically opposite angles, perpendicular lines, and the criteria and properties of parallel lines.
Year 7	2nd term	real number	Concepts of square roots, cube roots, irrational numbers, and real numbers.
Year 7	2nd term	Cartesian coordinate system	Ordered pairs, construction of a coordinate system, and representation of point locations.
Year 7	2nd term	System of linear equations in two variables	Substitution method, elimination method (addition/subtraction), and applications of systems of linear equations in two variables.
Year 7	2nd term	Inequalities and Systems of Inequalities	Properties of inequalities, methods for solving linear inequalities (and systems thereof), and representation on the number line.
Year 7	2nd term	Data Collection and Organization	Sample survey, complete enumeration, histogram, line chart.

In figure 3, a set of documents of a specific mathematics topic are collected. A classifier is trained in the first layer to classify the topic into sub-topics. In the second layer, the lexicon-based machine learning classifier is trained with the user defined mathematics sub-topics, its synonyms and the lexicon words of the sub-topic, and their neighbor words to build the multi-layer classifier to classify the document into the topic.

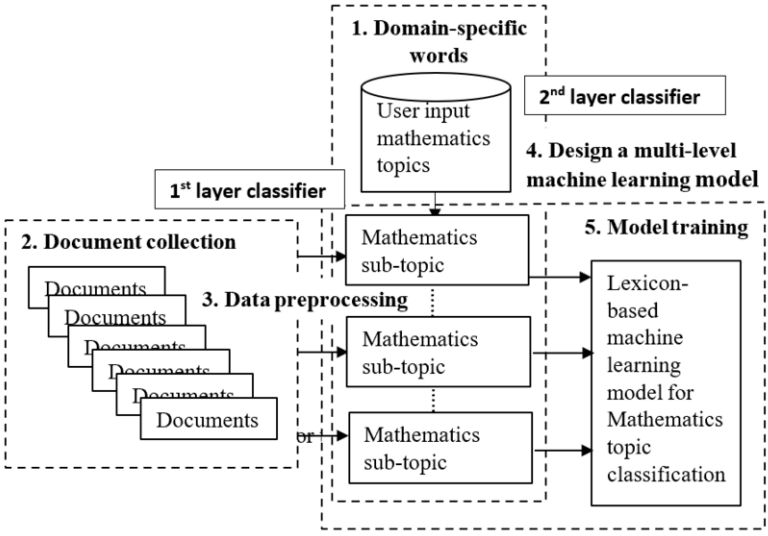


Figure 3. A Conceptual Design of the lexicon-based machine learning model.

To design the second level of the lexicon-based machine learning model, the sub-topic name, its synonyms and lexicon words, and the lexicon words' neighbor words in the document are used for training the model to classify the topic and its semantics meaning. The types of machine learning models can be used including rule-based systems, support vector machines, naïve bayes, random forest, BERT, RoBERTa, RNN, LSTMs, etc. The output of the model are the classified semantic context items such as topic name, learning objects types, year of study, level of difficulty, and some other context items. In the example of the "rational number" topic, the outputs are the "rational number", "a theory, a concept or an example", "Year 7 1st term or Year 7 2nd term", "easy or difficulty", etc.

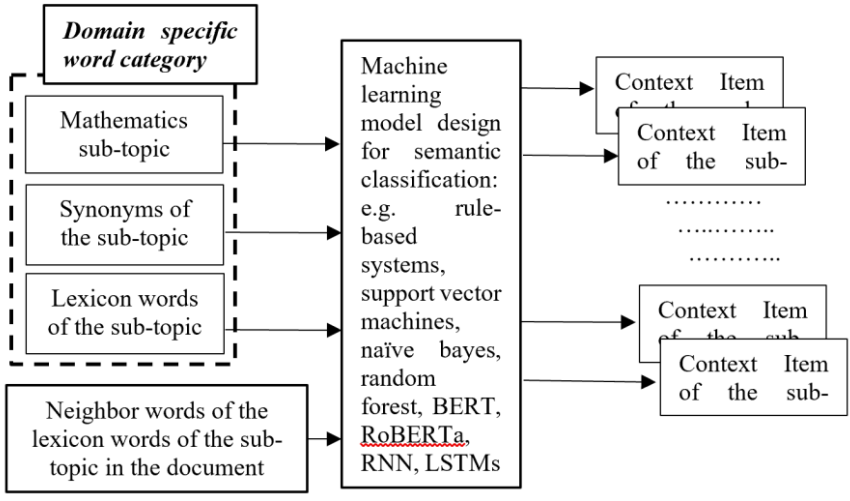


Figure 4. A Conceptual Design of the second layer of the lexicon-based machine learning model.

5. Discussion

To implement the two layers topic detection classifier, different machine learning methods such as rule-based systems, support vector machines, naïve bayes, random forest, BERT, RoBERTa, RNN, LSTMs [15,16,17] can be used. Each of them has its strengths and weaknesses. The rule-based system has high accuracy for a specific domain without training dataset for training. However, it cannot understand the context of the text for topic classification that may result in error. To train a better model, huge of data is required. Support vector machines can handle high-dimensional text data. However, it consumes lots of memory and time resources, and no probabilistic output. Naïve Bayes is a probabilistic classifier with less memory and time resources for training. However, it cannot classify the topic with a more complex relationships of the text. Random Forest can overcome it to learn a more complex relationships of the text in a topic. However, it requires significant time and effort to label a training dataset. Therefore, the supervised deep learning model such as BERT, RoBERTa, RNN, LSTMs are some other more advanced models to understand the semantic meaning, sequence, and context of words for topic classification. However, it consumes lots of time and resources to train and deploy the model though the accuracy can improve, and it can solve the complex text relationships. Therefore, developers can base on its memory, time and resource costs to select the appropriate method for context-based content analysis.

6. Conclusion

In summary, this research identified the limitations and challenges of the existing content analysis. A novel approach of lexicon-based machine learning method using semi-human and semi-machine learning methods was proposed to overcome the limitations. Table 2 summarized how our lexicon-based machine learning to be used to overcome the existing limitations of understanding on domain-specific language, colloquialisms and slang words, synonyms, and human contextual word meanings for text classification. The novelty of this paper is to develop a conceptual framework of the novel lexicon-based machine learning text classification model for text classification with semantic exploring on its classification. It makes the traditional text classification from content based to context based, and to exploration based. Other than text classification, this model can further to develop for finding out the semantic meaning of text for constructing the text summarization. The limitation of this study is the lack of experiments to provide a case study on how to develop the lexicon-based machine learning model for text classification. In future work, researchers will select and evaluate difference machine learning model, develop the lexical-based machine learning model, and run experiments to compare the proposed one with the traditional one for text classification.

Table 2. A Summary of how lexicon-based machine learning to be used.

The challenges of content analysis	Our proposed lexicon-based machine learning method
Domain-specific language	Construct a repository for user-defined domain-specific words
Colloquialisms and slang words	Construct a vocabulary repository to remove the colloquialisms and slang words in the document.
Synonyms problem	Use machine learning method such as fast text built in function to find the similar words of the user-defined sentiments
Human context word meaning	Use user-defined domain-specific words and synonyms to build the lexicon-based machine learning classifier

References

[1] Lau ASM. Implementation of an onto-wiki toolkit using web services to improve the efficiency and effectiveness of medical ontology co-authoring and analysis. Informatics for Health and Social Care. 2009:34(1):73 – 80.

[2] Tsui E, Wang WM, Cheung CF, and Lau ASM. A Concept-Relationship Acquisition and Inference Approach for Hierarchical Taxonomy Construction from Tags. Information Processing & Management. 2010;46(1), 44-57.

[3] Ahmed, A.Z. and Díaz, M.R. A Methodology for Machine-Learning Content Analysis to Define the Key Labels in the Titles of Online Customer Reviews with the Rating Evaluation. Sustainability. 2022;14(15): 9183. https://doi.org/10.3390/su14159183

[4] Srivastava, R., Bharti, P.K., and Verma, P. Comparative Analysis of Lexicon and Machine Learning Approach for Sentiment Analysis. International Journal of Advanced Computer Science and Applications. 2022;13(3). http://doi.org/10.14569/IJACSA.2022.0130312

[5] Hamed, S., Ezzat, M., and Hefny, H. A Review of Sentiment Analysis Techniques. International Journal of Computer Applications. 2020;176(37):20-24. http://doi.org:10.5120/ijca2020920480

[6] Veen, A.M. and Bleich, E. The advantages of lexicon-based sentiment analysis in an age of machine learning. PLoS One. 2025; 20(1):e0313092. http://doi.org/10.1371/journal.pone.0313092

[7] Mukhtar, N., Kha, M.A., and Chiragh, N. Lexicon-based approach outperforms Supervised Machine Learning approach for Urdu Sentiment Analysis in multiple domains. Telematics and Informatics. 2018:35(8): 2173-2183. http://doi.org:10.1016/j.tele.2018.08.003

[8] Kamiri, J. and Mariga, G. Research Methods in Machine Learning: A Content Analysis. International Journal of Computer and Information Technology. 2021;10(2). http://doi.org/10.24203/ijcit.v10i2.79

[9] Alsaffar A, Omar N. Integrating a Lexicon Based Approach and K Nearest Neighbour for Malay Sentiment Analysis. Journal of Computer Science. 2015; 11(4):639-644. doi:10.3844/jcssp.2015.639.644.

[10] Outahajala M, Rosso P. Using a Small Lexicon with CRFs Confidence Measure to Improve POS Tagging Accuracy. Proceedings of the Tenth International Conference on Language Resources and Evaluation; European Language Resources Association, c2016.p.4311-4315. doi:10.63317/2wg6j6reunrq.

[11] Koerich AL, Sabourin R, Suen C. Lexicon-Driven HMM Decoding for Large Vocabulary Handwriting Recognition with Multiple Character Models. International Journal on Document Analysis and Recognition. 2003;6(2):126-144. doi:10.1007/s10032-003-0113-0.

[12] Yelmen I, Gunes A, Zontul M. Multi-Class Document Classification Using Lexical Ontology-Based Deep Learning. Applied Sciences. 2023; 13(10):6139. doi.org/10.3390/app13106139.

[13] Mukherjee P, Leroy G, Kauchak D. Using Lexical Chains to Identify Text Difficulty: A Corpus Statistics and Classification Study. IEEE J Biomed Health Inform. 2019 Sep;23(5):2164-2173. doi: 10.1109/JBHI.2018.2885465.

[14] Anthony P, Zhou J. Leveraging Transformer with Self-Attention for Multi-Label Emotion Classification in Crisis Tweets. Informatics. 2025; 12(4):114. https://doi.org/10.3390/informatics12040114.

[15] Cheung L, Wang Y, Lau ASM, Chan RMC. Using a Novel Clustered 3D-CNN Model for Improving Crop Future Price Prediction. Knowledge-based System. 2023. https://doi.org 10.1016/j.knosys.2022.110133.

[16] Chen J, Fan Z, Liu Y, Wang Y, Xu Z, Zhao W, Lau ASM. "Using a Novel Multi-stage of TF-IDF, TextRank and KeyBert for Market Campaign Analysis. Frontier in Artificial Intelligence and Applications. 2026;424:292-314, https://doi.org/10.3233/FAIA260597.

[17] Wang B, Wen Y, Chen D, Yang X, Dou M, Zhong D, Lau ASM. A Credit Risk Chatbot by Using TF-IDF, XGBoost, and Random Forest for Improving Credit Risk Measurement. Frontier in Artificial Intelligence and Applications. 2025;424:348-363, https://doi.org/10.3233/FAIA260600.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261107

FinSight: An Evidence-First Financial Analysis Chatbot with Explainable Query Intelligence

Ruiqi XIE [a], Jiachang ZHANG [a], Mingxi SHI [a], Shiyi LI [a], Mingyuan GAO [a], Zihao YIN [a] and Adela S.M. LAU [b, 1]

[a] *Department of Statistics and Actuarial Science, School of Computing and Data Science, The University of Hong Kong, China*

[b] *Data Science Lab, Department of Statistics and Actuarial Science, School of Computing and Data Science, The University of Hong Kong, China*

ORCiD ID: Ruiqi Xie https://orcid.org/0009-0000-9585-8525
Jiachang Zhang https://orcid.org/0009-0007-2822-5657
Mingxi Shi https://orcid.org/0009-0003-8557-9620
Shiyi Li https://orcid.org/0009-0003-2876-1856
Mingyuan Gao https://orcid.org/0009-0005-7074-2637
Zihao Yin https://orcid.org/0009-0007-7444-8384
Adela S.M. Lau https://orcid.org/0000-0001-5918-8309

Abstract. Financial question answering is difficult because user queries often combine ambiguous intent, entity aliases, temporal scope, market data, news narratives, and advice-seeking language. Fluent responses become risky when the resolved entity, selected data sources, and evidence gaps remain hidden. This paper presents FinSight, an evidence-first financial analysis chatbot for questions about China's financial markets. The system separates evidence production from final-response generation. Its Query Intelligence backend converts a natural-language query into explicit natural-language understanding (NLU) and retrieval artifacts, including the normalized query, product type, intents, resolved entities, source plan, retrieved documents, structured records, evidence coverage, warnings, ranking traces, and analysis summary. Downstream modules provide frontend rendering, numerical analysis, document sentiment analysis, constrained LLM-based summarization, and follow-up-question prediction. In a local evaluation of 10,000 query-style cases, including 6,236 finance-domain and 3,764 out-of-scope queries in Chinese and English, FinSight achieved finance-domain recall of 0.975, out-of-scope rejection accuracy of 0.9729, intent micro-F1 of 0.9407, topic micro-F1 of 0.9208, and retrieval Recall@10 of 0.9664. In addition, all contract checks passed in a 32-case end-to-end fuzz matrix. Typed evidence handoffs allow readers to trace each generated claim to a retrieved item or explicit warning rather than treating the final response as unsupported output.

Keywords. financial question answering, query intelligence, evidence retrieval, sentiment analysis, large language models, decision support

[1] Corresponding Author: Adela S.M. Lau, Data Science Lab, Department of Statistics and Actuarial Science, School of Computing and Data Science, Run Run Shaw Building, The University of Hong Kong, Pokfulam Road, Hong Kong; E-mail: adelalau@hku.hk.

1. Introduction

Retail investors increasingly ask broad natural-language questions such as whether a stock is worth holding, why a company's share price fell on a given trading day, or how a macroeconomic policy may affect a sector. Such questions are challenging because they require entity disambiguation, financial intent detection, source selection, numerical evidence, textual interpretation, and risk-aware communication. Although some researchers developed prediction models using the numerical data [1] and textual information]2] for investment advice, a fluent system may produce a convincing response while obscuring the entity it resolved, the sources it used, and the limitations of its evidence. For financial questions, the system should expose the entity, intent, risk flag, and planned evidence before presenting quantitative results or explanatory text. The efficient market hypothesis and the adaptive markets hypothesis describe how public information, investor behavior, and market regimes interact, motivating caution about deterministic advice [3][4]. A financial question-answering system should therefore ground its responses in visible evidence rather than unsupported generation.

FinSight implements this separation across the chatbot frontend, the natural-language understanding (NLU) and retrieval backend, the numerical and text-analysis modules, and the large language model (LLM) summarizer. The backend creates nlu_result and retrieval_result objects that record normalized query, product type, intents, topics, resolved entities, missing slots, risk flags, source plan, retrieved documents, structured data, coverage, warnings, and analysis summary. The LLM module then consumes compact evidence and returns validated JSON containing answer text, key points, evidence identifiers, limitations, risk disclaimer, and follow-up questions.

This paper makes three contributions. First, it introduces an evidence-contract architecture in which a classical, inspectable Query Intelligence layer produces typed records before generation, and downstream modules consume those records without re-inferring the query intent or source plan. Second, it provides a ready-to-run bilingual implementation for questions about China's financial markets and evaluates its explainability through 10,000 query-style cases and a 32-case end-to-end contract test. Third, it demonstrates deterministic behavior in the core evidence-processing layer for fixed inputs and runtime assets, together with explicit coverage warnings and abstention for unsupported inputs.

2. Related Work

Financial decision support must distinguish observed evidence from predictions and advice. The efficient market hypothesis frames price formation around available information, whereas the adaptive markets hypothesis describes changing behavior across market regimes [3][4]. Quantitative finance further emphasizes risk factors, fundamentals, and market uncertainty [5][6]. These ideas motivate a cautious interface for FinSight: the system may summarize market, fundamental, and textual signals, but it must not present a buy-or-sell decision as an established fact.

The retrieval layer follows classical information-retrieval principles. Information needs must be represented, matched to documents, and evaluated through ranked relevance [7]. BM25 and probabilistic ranking remain strong, interpretable baselines for sparse document retrieval [8]. Learning-to-rank methods and metrics such as normalized discounted cumulative gain (NDCG) formalize the goal of placing more relevant evidence higher in the result list [9][10]. FinSight retains the source plan, candidate

counts, ranking scores, evidence identifiers, and warnings rather than preserving only the answer-facing output.

The NLU layer uses classical, explainable methods as its backbone. Conditional random fields support sequence labeling and entity-boundary detection [11]. Linear classifiers and support-vector methods remain effective for sparse text classification [12], whereas ensemble methods can capture nonlinear feature interactions while retaining inspectable behavior [12][14]. Because an incorrect entity or product type can mislead every downstream module, this layer is deliberately kept auditable.

Research on financial text shows why retrieved news and announcements should be analyzed separately from numerical data. Media content can affect investor sentiment and market movements [15], and financial language differs from general-domain text, motivating domain-specific lexicons and models [16]. FinBERT-style models adapt language pretraining to financial sentiment analysis [17][18]. FinSight applies this capability only after the retrieval layer has selected relevant evidence.

Retrieval-augmented generation (RAG) improves generated text by conditioning it on retrieved evidence [19]. FinSight uses RAG in a narrower, contract-driven manner: Query Intelligence performs source planning and retrieval, and the LLM summarizes only the compact evidence package it receives. JSON Schema, Pydantic-based validation, and typed web APIs support predictable contracts among the backend, LLM module, and frontend [20][21][22].

Recent financial-domain language models improve generation quality but often keep query understanding, entity resolution, source selection, and evidence grounding implicit within the model. Financial LLMs such as BloombergGPT [23], FinGPT [24], PIXIU/FinMA [25], and the Chinese model DISC-FinLLM [26] are pretrained or instruction-tuned on financial corpora. Financial question-answering datasets such as FinQA [27] and ConvFinQA [28] target numerical reasoning over reports, while surveys and holistic benchmarks document the field's rapid development [29][30]. Two studies published in 2025 address related RAG settings: FinDER focuses on realistic analyst questions containing ambiguity and financial terminology [31], whereas FinS-Pilot evaluates online financial RAG systems that combine live API data with text under intent classification [32]. FinSight differs in emphasis by performing intent detection, entity linking, and source planning in a separate, inspectable stage rather than within the generative pass. The numerical, sentiment, and LLM modules then consume that stage's output, cite its evidence, and respect its risk flags. The principal contribution is therefore the enforced contract between layers rather than a new financial foundation model.

3. System Architecture and Methods

Figure 1 shows the architecture, and Table 1 lists each module's main inputs and outputs. The frontend submits a user query to a FastAPI-based service. The backend first normalizes the query and classifies product type, intent, topic, and question style. Entity resolution combines aliases, exact matching, runtime entity tables, and typo-tolerant linking. The source planner selects sources such as market APIs, news, announcements, industry data, fundamentals, macro indicators, and local document retrieval. Retrieval then builds source-specific queries, deduplicates results, ranks candidates, packages evidence, and produces coverage flags and warnings.

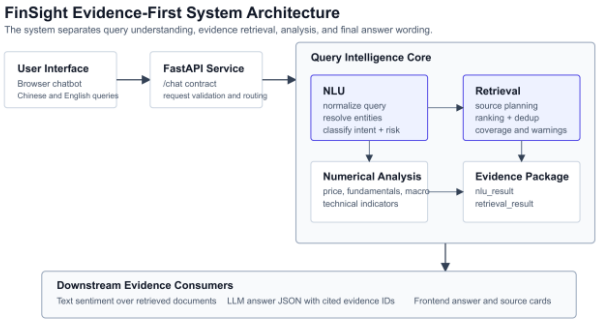


Figure 1. Query Intelligence produces explicit evidence artifacts; downstream modules perform numerical analysis, text analysis, LLM summary, and frontend rendering.

Table 1. Module-level data science design.

Module	Input artifacts	Main output artifacts
Frontend	Raw query, API response, language context	Chat answer, source cards, key points, disclaimer
NLU & Retrieval	Query text, aliases, runtime entities, documents, source registry	nlu_result, retrieval_result, ranked evidence, warnings
Numerical analysis	Market, fundamental, industry, macro rows	Returns, indicators, data-readiness, analysis_summary
Text analysis	Retrieved documents and NLU constraints	Evidence-linked sentiment items and excerpts
LLM summary	Compact evidence JSON	Validated answer JSON and follow-up questions

The numerical module consumes structured rows from live providers and shipped runtime data. It standardizes symbols, dates, provider names, query parameters, and numeric fields. When sufficient historical data exists, it computes descriptive technical indicators such as multi-window returns, MA5, MA20, RSI(14), MACD, volatility, Bollinger bands, and trend signals, following standard technical analysis definitions [33]. These indicators are treated as evidence summaries, not forecasts. Data-readiness flags tell downstream modules whether market, fundamental, macro, news, and technical evidence are available.

The text-analysis module consumes only documents already selected by retrieval. It filters by entity, product type, intent, source type, and risk flags before sentence segmentation and FinBERT-based sentiment classification using `yiyanghkust/finbert-tone-chinese` for Chinese and `ProsusAI/finbert` for English. The output is a document-level sentiment item that preserves evidence identifier, source, title, relevant excerpt, label, confidence, and rank score, keeping each sentiment claim linked to a retrieved document.

The LLM module converts compact evidence into frontend-ready JSON. It removes debug fields and lengthy raw records, calls a chat-completions-compatible API, validates the returned JSON structure, attaches the required risk disclaimer, and removes any statement unsupported by the evidence package. The same module predicts follow-up questions, but these predictions are workflow suggestions rather than financial signals.

The implementation can use a DeepSeek-compatible endpoint by default; however, the architecture remains provider-agnostic because the frontend and backend depend on a chat-completions API contract rather than a proprietary model interface.

4. Implementation and Data

FinSight is implemented as a Python project comprising a FastAPI service, a browser-based chatbot launcher, typed contracts, NLU and retrieval packages, a sentiment package, training scripts, tests, schemas, and runtime assets. The runtime data include entity master tables, alias tables, and local document fallbacks. Live providers include Tushare, AKShare, efinance, Cninfo-style announcement feeds, and optional PostgreSQL retrieval. Separately, the training pipeline uses documented public datasets to support reproducible model development. Large public raw datasets and generated training assets are documented by source rather than embedded in the repository.

Pre-processing is specific to each module. NLU normalizes punctuation, aliases, multilingual terms, and financial product mentions. Retrieval standardizes source names, timestamps, evidence identifiers, ranking features, and missing-data warnings. Numerical analysis converts provider payloads into tidy structured records and quality flags [34]. Text analysis performs language detection, source filtering, entity relevance checks, and sentence-level processing. LLM evidence packing removes long raw provider payloads and keeps only compact facts cited by evidence identifier.

The project separates source code from generated artifacts. Classical models are trained from manifest-driven public datasets with scikit-learn [35], and selected runtime assets are included to provide a ready-to-run implementation. Generated model binaries are reproducible artifacts created by the training scripts rather than scientific results; accordingly, this paper describes the data sources and training procedure without treating local caches as evaluation outputs.

5. Evaluation and Results

The 10,000-case evaluation set comprises twelve buckets covering public financial question answering, synthetically generated financial variants, public out-of-scope text, conversational out-of-scope dialogue, retrieval-specific queries, and small must-pass safety sets. Public buckets are sampled from Chinese- and English-language financial and general-domain corpora. Generated buckets are templated from real entities and intents in the runtime tables and are manually spot-checked. Although the must-pass buckets are small, they are mandatory: failure on any must-pass case causes the evaluation to fail regardless of the aggregate metrics. Table 2 summarizes the composition.

Table 2. Bucket counts in the 10,000-case Query Intelligence regression set.

Bucket	Cases	Purpose
Public finance	3,983	Real finance QA, Chinese and English
Public out-of-scope	2,743	General-domain text
Finance retrieval	1,397	Retrieval-specific finance queries
Public out-of-scope dialogue	1,000	Conversational out-of-scope queries
Generated finance: market movement	713	Synthetic “why did X move” queries
Generated finance: fact lookup	108	Synthetic fact-lookup queries

Bucket	Cases	Purpose
Must-pass out-of-scope	21	Must reject confidently
Generated finance: comparison	16	Synthetic comparison queries
Generated finance: macro policy	13	Synthetic macro-policy queries
Must-pass finance: open evaluation	2	Must answer with partial evidence
Must-pass finance: unknown entity	2	Must trigger clarification
Generated finance: clarification	2	Must flag ambiguity

Table 3. Summary of the Query Intelligence evaluation results.

Evaluation item	Observed result
Evaluation set	10,000 cases: 6,236 finance, 3,764 out-of-scope, 5,865 Chinese, and 4,135 English
Routing and NLU	Finance recall 0.975; out-of-scope rejection 0.9729; product and question-style accuracy 1.0; intent/topic micro-F1 0.9407/0.9208
Source planning and retrieval	Source-plan Hit@5 1.0; source-plan Recall@5 0.9957; Recall@10 0.9664; MRR@10 0.889; NDCG@10 0.8872
Boundary behavior	Out-of-scope retrieval abstention 0.986; clarification recall 0.75
Complementary checks	32/32 fuzz cases, 361/361 contract checks, frontend health, and bilingual chat tests passed

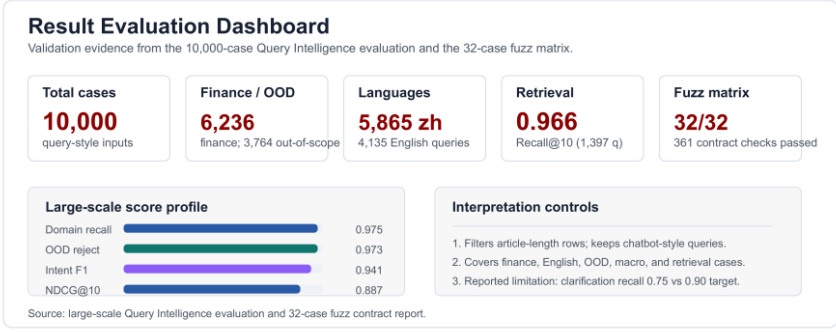


Figure 2. The dashboard summarizes the 10,000-case Query Intelligence evaluation, complementary fuzz tests, retrieval ranking, and answer-safety controls.

The main regression evidence is a 10,000-case Query Intelligence evaluation, as shown in Table 3. The set combines public and generated query-style examples, filters out article-length rows, and measures domain routing, labeling, source planning, retrieval ranking, and out-of-scope abstention. Ground truth comes from public dataset annotations and the templated fields of the synthetic buckets, while retrieval relevance uses a source registry mapping each query to its expected sources.

The 32-case fuzz matrix provides a complementary end-to-end contract test. The hand-curated cases cover colloquial phrasing, typographical errors, inputs requiring clarification, macroeconomic-policy questions, and English-language queries. Each case is evaluated using automatic checks of nlu_result and retrieval_result schema validity, product type, question style, entity symbols, temporal scope, source-plan membership, risk flags, warnings, executed sources, and per-source coverage; a case passes only when

every applicable check succeeds. All 32 cases passed all 361 checks, including every schema-validity check. For fixed inputs and runtime assets, the core evidence-processing stages are deterministic, whereas end-to-end latency is dominated by optional external provider and LLM calls. Planned-source coverage was high, with source-plan Hit@5 of 1.0 and source-plan Recall@5 of 0.9957, indicating that expected sources were usually scheduled before retrieval.

A representative Chinese-language case, translated here, is "Why did Kweichow Moutai fall today, and is it still worth holding?" The normalized query resolved to Kweichow Moutai, symbol 600519.SH, through exact alias matching. The NLU module classified the product type as stock, detected market-explanation, hold-judgment, and buy-sell-timing intents, and set an advice-like risk flag. The source plan selected news, market, fundamental, industry, and announcement sources. Retrieval returned ranked evidence from the news, market, fundamental, and industry sources and retained an announcement-missing warning for the recent window. The response was therefore restricted to available evidence and explicitly identified the missing announcement evidence.

The numerical result included a market row with trade date, close price, daily percentage change, volume, and amount, plus fundamental evidence for revenue, net profit, ROE, PE, and PB, and industry evidence for sector change, valuation, and turnover. These fields support descriptive interpretation rather than a deterministic investment recommendation.

To make the evidence contract concrete, the same case produces compact artifacts that travel end to end. The nlu_result carries product_type=stock, entities=[600519.SH], intents=[market_explanation, hold_judgment, buy_sell_timing], and risk_flags=[investment_advice_like]. The retrieval_result lists executed sources [news, market, fundamental, industry], an announcement_missing warning, and an analysis_summary. The final answer_generation JSON cites these by identifier, e.g. evidence_ids=["news#3","market#1"], and always includes a risk disclaimer, so a reader can audit which retrieved item supports each claim.

A second, contrasting case demonstrates how the contract adapts to the query type. For the macroeconomic question "How would a reserve-requirement-ratio cut affect bank stocks?", the nlu_result sets intents=[macro_impact, sector_analysis] without resolving a single stock. The source plan selects macroeconomic indicators, policy events, and industry data rather than a market API, and the answer cites the policy and industry evidence IDs while stating that the effect is conditional.

Figure 2 consolidates these metrics in a single evaluation dashboard. When run locally through the browser-based chatbot launcher, the system returned answer text, key points, evidence cards, and risk disclaimers for both Chinese and English questions; English input produced English output. The LLM handoff also returned three follow-up-question predictions with scores and rationales to guide further exploration.

For context, BEIR includes the FiQA-2018 retrieval task [36], and T2Ranking provides a large-scale Chinese passage-ranking benchmark [37]. Their published results use different datasets, relevance definitions, and evaluation protocols from those used for FinSight. Consequently, their scores should be treated as contextual reference points rather than direct performance comparisons. FinSight's evaluation additionally measures source-plan accuracy, evidence coverage, abstention, and risk controls that these general retrieval benchmarks were not designed to assess.

5.1. Failure Analysis

Error analysis identified 39 finance-domain false negatives in the routing-evaluated subset, consistent with the finance-domain recall of 0.975. It also identified intent and topic disagreements selected for root-cause inspection and one missed clarification among the four cases requiring clarification, corresponding to a recall of 0.75. Three representative failures illustrate the principal causes.

(a) False out-of-scope classification. A Chinese news excerpt stating that Netflix, Amazon, and Pandora use artificial intelligence had a financial-domain gold label in the public dataset but was classified as out of scope. Product-type classification depends partly on entity coverage, and these foreign streaming firms are absent from the China-market alias set. Expanding the alias table or adding a foreign-equity pass-through rule could mitigate this error.

(b) Intent overprediction. The cross-lingual query "Why did Guoxin Health rally after the drawdown?" produced both market_explanation and risk_analysis when only the former was expected. This result suggests that the classifier is overly permissive for advice-related, code-switched language; topic micro-precision was 0.8549, compared with micro-recall of 0.9977.

(c) Missed clarification. The query "Is Xingchen Technology worth attention?" contains an unregistered alias. The system set the investment_advice_like flag but did not set clarification_required or missing_entity. Missing-slot detection currently triggers for absent time or amount expressions but not when entity resolution returns no match, which explains the observed clarification-recall gap.

6. Discussion

The design can be compared with three alternative architectures. In an LLM-only chatbot, entity resolution and source selection remain largely implicit within the model. A vector-only RAG system offers limited control over product-type routing and may not expose a source-planning rationale. A quantitative dashboard provides structured metrics but does not, by itself, model user intent, document sentiment, or bilingual explanation. FinSight records these decisions in an explicit stage, making the resolved entity, source schedule, coverage gaps, and scope of each claim visible. The 10,000-case regression evaluation and the 32-case matrix assess these properties in addition to the quality of the final prose.

Several limitations remain. Retrieval depends on external data providers, including Tushare, AKShare, and announcement feeds, which may impose rate limits or return incomplete records. The system reports coverage warnings for such gaps but cannot reconstruct data omitted by a provider. Ambiguous or unregistered company names also remain challenging: clarification recall was 0.75 against a target of 0.90 because missing-slot detection does not yet trigger when entity resolution returns no match. The numerical module is descriptive; it reports indicators and, for speculative or future-oriented queries such as multiyear price forecasts, returns evidence and uncertainty warnings rather than point predictions. Document-level sentiment is evidence-linked, but time-aware aggregation remains future work. The LLM path depends on prompt adherence and JSON validation; fallback procedures reduce the risk of invalid output but cannot create missing evidence. Finally, the 10,000-case set combines public and synthetic examples and may not represent live traffic, so the results require confirmation in deployment.

7. Conclusion

This paper presented FinSight, an evidence-first financial analysis chatbot with explainable Query Intelligence. The system converts natural-language financial questions into explicit NLU and retrieval artifacts, then performs numerical analysis, document sentiment analysis, and constrained LLM-based summarization over those artifacts. Evaluation on 10,000 query-style cases and a 32-case end-to-end fuzz matrix demonstrated bilingual browser-chat responses with evidence cards and risk disclaimers while exposing the entities, intents, sources, warnings, and evidence identifiers underlying each response. Future work will monitor live data sources, improve clarification recall, add time-aware sentiment aggregation, and evaluate the system on larger unseen question sets.

References

[1] Cheung L, Wang Y, Lau ASM, Chan RMC. Using a Novel Clustered 3D-CNN Model for Improving Crop Future Price Prediction. Knowledge-based System. 2023. https://doi.org/10.1016/j.knosys.2022.110133.
[2] Nie RX, Liu ZY, Yu WY, Zhan RQ, Lau ASM. The Application of Artificial Intelligence in Financial Prediction. Frontier in Artificial Intelligence and Applications. 2025;412, https://doi.org/10.3233/FAIA250732.
[3] Fama EF. Efficient capital markets: a review of theory and empirical work. J Finance. 1970;25(2):383-417. doi: 10.1111/j.1540-6261.1970.tb00518.x.
[4] Lo AW. The adaptive markets hypothesis: market efficiency from an evolutionary perspective. J Portfolio Manag. 2004;30(5):15-29. doi: 10.3905/jpm.2004.442611.
[5] Fama EF, French KR. Common risk factors in the returns on stocks and bonds. J Financ Econ. 1993;33(1):3-56. doi: 10.1016/0304-405X(93)90023-5.
[6] Hull JC. Risk management and financial institutions. 5th ed. Hoboken (NJ): Wiley; 2018.
[7] Manning CD, Raghavan P, Schütze H. Introduction to information retrieval. Cambridge: Cambridge University Press; 2008. doi: 10.1017/CBO9780511809071.
[8] Robertson S, Zaragoza H. The probabilistic relevance framework: BM25 and beyond. Found Trends Inf Retr. 2009;3(4):333-389. doi: 10.1561/1500000019.
[9] Burges CJC, Shaked T, Renshaw E, Lazier A, Deeds M, Hamilton N, et al. Learning to rank using gradient descent. In: Proceedings of the 22nd International Conference on Machine Learning; 2005. p. 89-96. doi: 10.1145/1102351.1102363.
[10] Järvelin K, Kekäläinen J. Cumulated gain-based evaluation of IR techniques. ACM Trans Inf Syst. 2002;20(4):422-446. doi: 10.1145/582415.582418.
[11] Lafferty J, McCallum A, Pereira FCN. Conditional random fields: probabilistic models for segmenting and labeling sequence data. In: Proceedings of the 18th International Conference on Machine Learning; 2001. p. 282-289.
[12] Cortes C, Vapnik V. Support-vector networks. Mach Learn. 1995;20(3):273-297. doi: 10.1007/BF00994018.
[13] Breiman L. Random forests. Mach Learn. 2001;45(1):5-32. doi: 10.1023/A:1010933404324.
[14] Friedman JH. Greedy function approximation: a gradient boosting machine. Ann Stat. 2001;29(5):1189-1232. doi: 10.1214/aos/1013203451.
[15] Tetlock PC. Giving content to investor sentiment: the role of media in the stock market. J Finance. 2007;62(3):1139-1168. doi: 10.1111/j.1540-6261.2007.01232.x.
[16] Loughran T, McDonald B. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. J Finance. 2011;66(1):35-65. doi: 10.1111/j.1540-6261.2010.01625.x.
[17] Araci D. FinBERT: financial sentiment analysis with pre-trained language models. arXiv [Preprint]. 2019. arXiv:1908.10063.
[18] Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of NAACL-HLT; 2019. p. 4171-4186. doi: 10.18653/v1/N19-1423.
[19] Lewis P, Perez E, Piktus A, Petroni F, Karpukhin V, Goyal N, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. Adv Neural Inf Process Syst. 2020;33:9459-9474.
[20] JSON Schema. JSON Schema: a media type for describing JSON documents [Internet]. 2022 [cited 2026 May 3]. Available from: https://json-schema.org/draft/2020-12/json-schema-core.html.
[21] FastAPI. FastAPI documentation [Internet]. [cited 2026 May 3]. Available from: https://fastapi.tiangolo.com/.

[22] Pydantic. Pydantic documentation [Internet]. [cited 2026 May 3]. Available from: https://docs.pydantic.dev/.
[23] Wu S, Irsoy O, Lu S, Dabravolski V, Dredze M, Gehrmann S, et al. BloombergGPT: a large language model for finance. arXiv [Preprint]. 2023. arXiv:2303.17564.
[24] Yang H, Liu XY, Wang CD. FinGPT: open-source financial large language models. arXiv [Preprint]. 2023. arXiv:2306.06031.
[25] Xie Q, Han W, Zhang X, Lai Y, Peng M, Lopez-Lira A, et al. PIXIU: a comprehensive benchmark, instruction dataset and large language model for finance. Adv Neural Inf Process Syst. 2023;36:33469-33484. doi: 10.52202/075280-1454.
[26] Chen W, Wang Q, Long Z, et al. DISC-FinLLM: a Chinese financial large language model based on multiple experts fine-tuning. arXiv [Preprint]. 2023. arXiv:2310.15205.
[27] Chen Z, Chen W, Smiley C, Shah S, Borova I, Langdon D, et al. FinQA: a dataset of numerical reasoning over financial data. In: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing; 2021. p. 3697-3711. doi: 10.18653/v1/2021.emnlp-main.300.
[28] Chen Z, Li S, Smiley C, Ma Z, Shah S, Wang W-Y. ConvFinQA: exploring the chain of numerical reasoning in conversational finance question answering. In: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing; 2022. p. 6279-6292. doi: 10.18653/v1/2022.emnlp-main.421.
[29] Li Y, Wang S, Ding H, Chen H. Large language models in finance: a survey. In: Proceedings of the 4th ACM International Conference on AI in Finance; 2023. p. 374-382. doi: 10.1145/3604237.3626869.
[30] Xie Q, Han W, Chen Z, Xiang R, Zhang X, He Y, et al. FinBen: a holistic financial benchmark for large language models. Adv Neural Inf Process Syst. 2024;37:95716-95743. doi: 10.52202/079017-3033.
[31] Choi C, Kwon J, Ha J, Choi H, Kim C, Lee Y, et al. FinDER: financial dataset for question answering and evaluating retrieval-augmented generation. In: Proceedings of the 6th ACM International Conference on AI in Finance; 2025. p. 638-646. doi: 10.1145/3768292.3770361.
[32] Wang F, Sun Y, Mao J, Xue W, Xu D. FinS-Pilot: a benchmark for online financial RAG system. In: Proceedings of the 34th ACM International Conference on Information and Knowledge Management; 2025. p. 6544-6548. doi: 10.1145/3746252.3761643.
[33] Murphy JJ. Technical analysis of the financial markets. New York (NY): New York Institute of Finance; 1999.
[34] Wickham H. Tidy data. J Stat Softw. 2014;59(10):1-23. doi: 10.18637/jss.v059.i10.
[35] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. J Mach Learn Res. 2011;12:2825-2830.
[36] Thakur N, Reimers N, Rücklé A, Srivastava A, Gurevych I. BEIR: a heterogeneous benchmark for zero-shot evaluation of information retrieval models. Adv Neural Inf Process Syst. 2021;34.
[37] Xie X, Dong Q, Wang B, Lv F, Yao T, Gan W, et al. T2Ranking: a large-scale Chinese benchmark for passage ranking. In: Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval; 2023. p. 2681-2690. doi: 10.1145/3539618.3591874.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261108

Supermarket CRM Intelligent Analysis System: A Data Mining Approach for Customer Segmentation, Churn Prediction, and Personalized Recommendation

Yanting GUO [a], Zixin WU [a], Yumeng WANG [a], Zhanhao LEI [a], Junyi GAO [a], Peizhen LYU[a], Boyu FENG [a], and Adela S.M. LAU[b,1]

[a] *Department of Statistics and Actuarial Science, School of Computing and Data Science, The University of Hong Kong, China*

[b] *Data Science Lab, Department of Statistics and Actuarial Science, School of Computing and Data Science, The University of Hong Kong, China*

ORCID ID:
Yanting Guo https://orcid.org/0000-0002-2047-7143
Zixin Wu https://orcid.org/0009-0001-7393-9049
Yumeng Wang https://orcid.org/0009-0004-5454-8263
Zhanhao Lei https://orcid.org/0009-0002-8280-9075
Junyi Gao https://orcid.org/0009-0007-0847-3194
Peizhen Lyu https://orcid.org/0009-0009-4675-8996
Boyu Feng https://orcid.org/0009-0004-4931-5705
Adela S.M. Lau https://orcid.org/0000-0001-5918-8309

Abstract. Escalating retail competition has driven grocery supermarkets to pivot from product-centric to customer-centric paradigms. Traditional Customer Relationship Management (CRM) paradigms rely heavily on heuristic or empirical experience, lacking systematic customer profiling, proactive churn risk identification, and automated personalized marketing. To address these limitations, this paper develops and implements an integrated CRM decision-support architecture powered by data mining techniques. Transitioning from conventional heuristic practices, this unified framework systematically enhances managerial decision-making by linking multi-dimensional customer segmentation, Random Forest-based churn prediction, a hybrid personalized recommendation engine, and data-driven smart recall strategies. Validated on a multi-constrained synthetic retail dataset embedded with industry-standard consumption dependencies (comprising 300 unique customer profiles and extensive transaction histories), the churn prediction module achieves an F1-score of 0.883 and an AUC of 0.943. Theoretical and practical insights are provided to demonstrate how brick-and-mortar retailers can leverage interpretable data pipelines to systematically curb customer attrition and maximize customer lifetime value.

[1] Corresponding Author: Adela S.M. Lau, Data Science Lab, Department of Statistics and Actuarial Science, Run Run Shaw Building, The University of Hong Kong, Pokfulam Road, Hong Kong; Email: adelalau@hku.hk

Keywords. Data Mining, Customer Relationship Management (CRM), Customer Segmentation, RFM Model, Random Forest, Collaborative Filtering, Association Rule Mining.

1.Introduction

Driven by rapid advancements and digitalization in the retail sector, competition among physical and omni-channel supermarkets has intensified exponentially. Consequently, traditional business models centered primarily on product turnover are rapidly shifting toward customer-centric paradigms, where understanding behavioral dynamics and maximizing customer retention have become core deterministic factors for sustained competitive advantage [1]. In this context, modern Customer Relationship Management (CRM) systems have emerged as an indispensable mechanism for elevating customer satisfaction, mitigating attrition, and executing refined operational strategies [2]. Within supermarket retail environments, robust CRM infrastructure enables enterprises to effectively profile high-value consumer segments, forecast latent churn risks, and deliver personalized marketing intervention campaigns [3]. Broadly, this integrated imperative aligns with the foundational theoretical consensus that deploying data-driven workflows is a crucial driver for modern transactional relationship optimization and retail intelligence transformation [4, 5].

However, conventional CRM frameworks deployed in conventional retail environments suffer from a heavy reliance on subjective empirical experience rather than systematic, data-driven insights. This reliance frequently manifests as ambiguous customer profiling, poorly targeted marketing, and lagged, reactive churn detection [6]. While modern supermarket transactional touchpoints generate massive volumes of purchase logs containing rich behavioral prints, extracting actionable business intelligence from these raw streams remains challenging without specialized, end-to-end analytical architectures.

To explicitly highlight our academic and practical contributions, this paper establishes an integrated CRM decision-support architecture. Unlike conventional retail CRM practices that operate in siloed configurations or rely heavily on intuitive heuristics, our unified architecture bridges the gap between raw data streams and closed-loop strategic interventions. To evaluate its managerial superiority, Table 1 contrasts the proposed integrated architecture against conventional CRM practices across key operational dimensions.

This study develops and implements an intelligent CRM analytical system anchored in robust data mining methodologies. The proposed system establishes a comprehensive analytical pipeline that seamlessly transforms raw transaction data into structured, actionable enterprise intelligence to facilitate data-driven consumer lifecycle management.

2. Operational Challenges in Retail CRM

Despite the extensive adoption of CRM concepts across the retail domain, modern supermarkets continue to confront several persistent operational bottlenecks. First, current CRM practices remain largely decoupled from rigorous, multi-dimensional data analysis. Decision-making processes concerning customer retention and targeted

promotion campaigns are predominantly guided by human intuition or static, hard-coded business rules. This lack of analytical rigor leads to coarse consumer profiling and inefficient capital and resource allocation [7].

Second, supermarkets lack automated, granular customer segmentation methodologies. In the absence of an integrated classification framework that dynamically evaluates customer lifecycle stages, lifetime value, and engagement depth, marketing strategies remain inherently undifferentiated. Consequently, premium loyal customers receive generic, low-ROI promotions, while dormant cohorts requiring substantial incentives are exposed to identical, ineffective touchpoints. To counteract this, contemporary empirical attempts utilize unsupervised clustering coupled with standard RFM metric partitioning to isolate broad customer groups [8]. However, these approaches consistently fail to dynamically map static statistical clusters onto granular, shifting consumer lifecycle progression paths.

Third, customer attrition management remains fundamentally reactive. Supermarkets typically recognize customer churn only after extended periods of total dormancy. At this advanced stage, reacquisition costs vastly exceed the financial outlays required for proactive, early-stage retention interventions [9]. The absence of early-warning mechanisms prevents operators from detecting subtle behavioral decay and deploying timely preventive actions.

Fourth, standard recommendation mechanics lack true personalization. Consumers are frequently bombarded with uniform promotional messages irrespective of their distinct transaction histories, preferences, or value segments. This mismatch results in depressed marketing conversion rates, low customer engagement, and substantial budget waste. Finally, user recall strategies for inactive segments lack empirical validation and optimization. Without an objective understanding of customer-specific churn drivers or custom-tailored incentive thresholds, re-engagement campaigns operate blindly, yielding poor return on investment (ROI).

Table 1. Operational Comparison between Conventional CRM and the Proposed Integrated Architecture.

Dimension	Conventional CRM Practice	Proposed Integrated Architecture	Managerial Impact / Improvement
Data Utilization	Siloed, static rule sets (e.g., rigid RFM thresholds)	Multi-dimensional engineered features with unsupervised clustering	Eliminates empirical bias; discovers organic, high-resolution consumer structures.
Customer Segmentation	Static, heuristic RFM partitioning with rigid thresholds	Multi-dimensional synthesis (RFM × Lifecycle × *K*-Means)	Captures hidden consumer migration paths and organic, granular behavioral sub-cohorts.
Churn Management	Reactive (detecting attrition only after extended dormancy)	Proactive predictive modeling (Random Forest risk scoring)	Shifts operations from reactive auditing to early-warning intervention, minimizing win-back costs.
Marketing Logistics	One-size-fits-all promotions, manual rule crafting	Three-layer hybrid recommendation fusion (Segment + CF + Apriori)	Maximizes cross-selling efficiency and budget ROI through precise item-customer matching.
Workflow Integration	Disparate tools requiring extensive manual data shuffling	Unified end-to-end processing pipeline deployed via Streamlit	Lowers technical barriers for store managers; automates message and voucher synthesis.

3. Related Work and Technological Paradigm Shifts

3.1. Limitations of Current Solutions

Conventional supermarket CRM infrastructures depend heavily on manual auditing and rudimentary rule-based tools, resulting in rigid and isolated analytical frameworks. Traditional Recency, Frequency, and Monetary (RFM) implementations utilize arbitrary, static thresholds that fail to adapt to evolving consumer behaviors or uncover latent purchasing cohorts, leading to overly rigid customer groupings. Furthermore, churn assessments typically depend on linear heuristics or standard logistic regression. These models assume linear relationships and fail to capture the highly complex, non-linear dependencies inherent in real-world retail behavior, thus yielding high false-negative rates in risk predictions [10]. From a promotional perspective, recommendation mechanisms remain one-size-fits-all, crafted manually without exploiting historical basket associations or collaborative signals. Finally, existing systems lack an integrated, production-ready pipeline, necessitating extensive manual data shuffling between disparate processing stages, which severely hinders scalability for large enterprise customer bases.

3.2. The Role of AI and Data Mining

Advanced artificial intelligence (AI) and data mining paradigms offer native solutions to these systemic limitations. Unsupervised machine learning algorithms, such as K-Means clustering, can autonomously partition vast behavioral spaces, uncovering organic consumer structures that escape rule-based sorting [11]. Concurrently, ensemble learning architectures, such as Random Forest, handle high-dimensional behavioral features natively to model non-linear feature interactions. This capability delivers highly accurate churn forecasting alongside interpretable feature importance rankings, allowing managers to identify the root drivers of customer attrition [12].

For personalized engagement, user-based and item-based collaborative filtering capture peer-group preference patterns to uncover latent product affinities, while market basket association rule mining isolates sequential purchasing dependencies to maximize cross-selling efficiency [13]. By embedding these algorithmic modules into a unified automated pipeline—spanning automated data preprocessing, real-time risk scoring, localized recommendation, and adaptive coupon synthesis—the system minimizes human intervention. When coupled with modular web visualization layers, these sophisticated analytics become fully accessible to operational stakeholders.

Recent advancements from 2023 to 2026 have increasingly focused on deep representation learning, such as Graph Neural Networks (GNNs) for capturing high-order topological relationships in transaction networks, and Large Language Models (LLMs) for conversational marketing. In the domain of predictive retention modeling, ensemble pipelines like XGBoost have been applied to capture sudden transactional drops [14]. Despite their high predictive accuracy, these deep frameworks exhibit prohibitive computational footprints, require continuous GPU infrastructure, and introduce severe black-box interpretability deficits. For small-to-medium physical supermarkets, a low-latency, interpretable, CPU-efficient architecture that cleanly synthesizes segmentation, prediction, and association remains a critical operational missing link, as standard frameworks typically address predictive churn and multi-layer

marketing logistics in isolation rather than within a unified, closed-loop analytical infrastructure [15].

4. Proposed System Architecture

4.1. System Design Overview

This work introduces an end-to-end intelligent CRM architecture that integrates diverse data mining techniques into a streamlined operational pipeline. The pipeline maps raw transaction streams onto actionable corporate initiatives through six functional layers: (1) Synthetic Data Generation & Ingestion, (2) Multi-dimensional Feature Engineering, (3) Hybrid Customer Segmentation, (4) Random Forest-driven Churn Prediction, (5) Three-layer Hybrid Recommendation, and (6) Adaptive Smart Recall.

The framework was validated utilizing a multi-constrained synthetic retail dataset designed to mirror authentic retail dynamics. The dataset captures 300 unique customer profiles, 50 distinct stock-keeping units (SKUs) across various product categories, and 3,365 transactional records, yielding a cumulative gross revenue of approximately RMB 224,302.60.

To ensure the synthetic dataset faithfully mirrors authentic retail dynamics rather than stochastic noise, several industry-specific operational constraints were mathematically embedded. First, inter-purchase intervals were sampled from random distributions calibrated to reflect realistic shopping frequencies across different customer segments (e.g., active customers purchase every 5–20 days on average, while less active cohorts exhibit longer gaps of 20–60 days). Second, item prices were manually calibrated to follow a realistic right-skewed distribution, ensuring that the majority of daily necessities remain low-cost (< RMB 30) while a long tail of premium goods extends to higher price points. Most importantly, prior association rules (e.g., historical co-purchasing probability vectors for complementary product pairs like toothpaste and toothbrushes) were strategically injected into the transaction generation process, providing a deterministic baseline for subsequent market basket analysis.

4.2. Customer Feature Construction and Preprocessing

Before execution of downstream clustering, raw purchase sequences are aggregated into high-order feature spaces. We implement data format standardization to regularize heterogeneous purchase dates into unified scalar values. Let $T_{current}$ denote the system assessment timestamp. The primary feature space d'i for customer i is formulated as follows:

$$R_i = T_{current} - max(t_i, m) \tag{1}$$

$$F_i = \sum I\left(Order_{i,m}\right) \tag{2}$$

$$M_i = \sum Amount_{i,m} \tag{3}$$

Where $I(\cdot)$ is the indicator function. To balance demographic and dynamic insights, we append secondary features: mean order value ($\mu_{M,i} = M_i/F_i$), coupon redemption rate (γ_i), and average feedback rating ($\overline{r_1}$).

4.3. Customer Segmentation Module

To balance rule-based business interpretability with algorithmic objectivity, the segmentation framework merges three distinct analytical vectors:

1) RFM Stratification: Customers are evaluated via normalized Recency, Frequency, and Monetary dimensions, dividing the base into five distinct tiers: High-Value Loyal (29.7%), Key Retention (10.0%), Potential Growth (16.0%), At-Risk (2.7%), and Lost & Inactive (41.7%) [14]. Figure 1 illustrates this distribution.

2) Customer Lifecycle Mapping: By analyzing historical inactivity windows and purchase frequency thresholds, cohorts are mapped into four lifecycle states: Churned (52.7%), Stable (29.7%), Active (16.7%), and Loyal (1.0%). Figure 2 presents this lifecycle distribution.

3) K-Means Clustering Pipeline: Using the six continuous behavioral features—recency, frequency, monetary value, mean order value, coupon redemption rate, and average feedback rating—the feature space is optimized via Silhouette Analysis. The K-Means objective function minimizes the total within-cluster variance:

$$\min_{C} \sum_{k=1}^{K} \sum_{x_i \in C_k} \|x_i - \mu_k\|^2 \tag{4}$$

where C_k represents the k-th cluster and μ_k is the corresponding centroid.

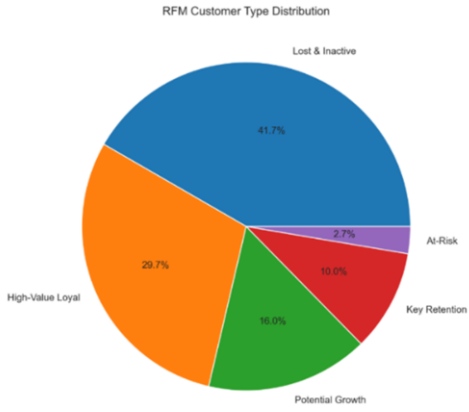


Figure 1. RFM customer type distribution: Five-tier segmentation of customer value

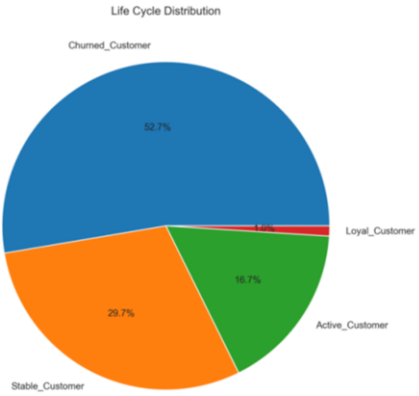


Figure 2. Customer lifecycle distribution: Four-stage classification of engagement status

To mathematically justify the selection of the optimal number of clusters K and eliminate empirical bias, a silhouette analysis was rigorously conducted for K ∈ [2, 10]. As illustrated in Figure. 3, the silhouette score achieves its global maximum (≈ 0.342) strictly at K = 2, followed by a monotonic decline as K increases. This numerical evidence demonstrates that partitioning the customer base into two highly distinct structural cohorts—namely, high-value loyal customers and low-value inactive customers—yields the most mathematically stable and managerially interpretable clustering configuration. The boundary thresholds for Recency, Frequency, and Monetary (RFM) variables were further calibrated based on three times the average purchasing cycle (μ) observed in the baseline data, consistent with established retail CRM practices. Silhouette analysis indicates K=2 as the mathematically optimal choice for this synthetic dataset, which reflects a highly polarized customer base (active vs. churned) rather than an inherent limitation of the methodology. In practice, more granular segments can be obtained by incorporating additional behavioral features or by relaxing the silhouette criterion.

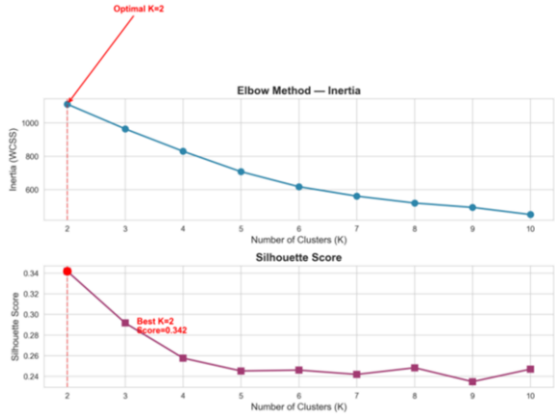


Figure 3. Silhouette coefficient analysis across varying cluster numbers K.

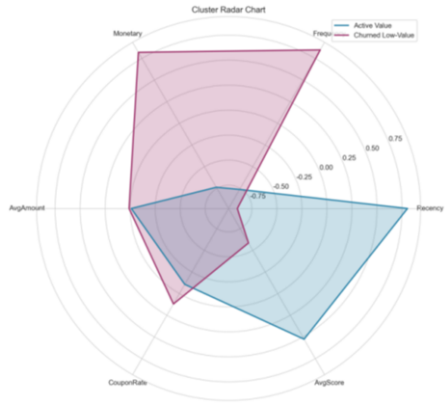


Figure 4. K-Means clustering radar chart: Comparative behavioral profiles of Active_Value_Customer vs. Churned_Low_Value_Customer clusters

4) Cross-Analysis Matrix: Overlapping the structural RFM tiers with lifecycle classifications reveals critical operational insights. Figure 5 presents this cross-analysis heatmap, revealing a dominant cohort of 124 customers residing precisely at the intersection of the Lost & Inactive and Churned states, highlighting a substantial target for win-back campaigns, while 77 High-Value Loyal users remain anchored in the Stable phase, representing the primary revenue engine.

Based on the cross heatmap results, the supermarket can directly deploy 35% deep-discount recall campaigns targeting these 124 lost & churn users, while arranging exclusive premium services for stable high-value loyal customers to lock long-term consumption. This actionable mapping explicitly bridges granular customer matrix cells with tailored enterprise resource allocations.

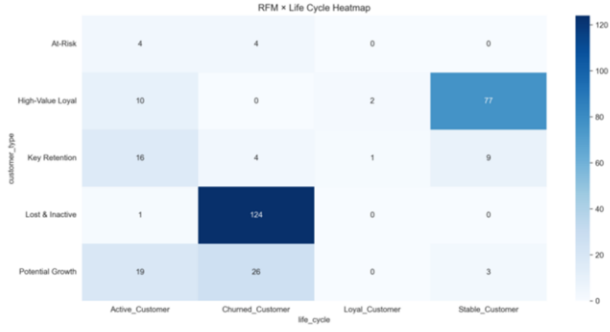


Figure 5. RFM and lifecycle cross-analysis heatmap: Customer count distribution across RFM tiers and lifecycle stages

4.4. Churn Prediction Module

The customer attrition risk engine utilizes a Random Forest framework, aggregating predictions across an ensemble of unpruned decision trees. To ensure unbiased evaluation under class imbalance, the dataset is partitioned using stratified sampling, guaranteeing that the proportion of churned versus active customers remains consistent between training and testing splits. Key hyperparameters, including the number of

estimators (n_estimators), maximum depth (max_depth), and minimum split samples (min_samples_split), are tuned via grid search. Given an ensemble of T individual trees, the predicted churn probability P_{churn} for a given instance x is formalized as:

$$P_{churn} = \frac{1}{T}\sum_{t=1}^{T} h_t(x) \tag{5}$$

Empirical feature importance extraction, shown in Figure 6, demonstrated that historical purchase frequency serves as the dominant predictive signal (importance score = 0.430), followed by cumulative monetary expenditure (0.253), average product rating (0.158), mean order value (0.077), consumer age (0.042), and coupon utilization frequency (0.039). These results empirically demonstrate that historical engagement behavior, rather than static demographic indicators, dictates retention probability.

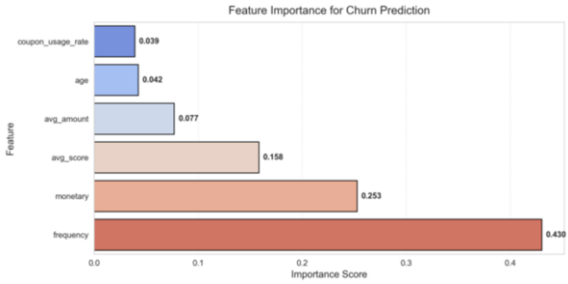


Figure 6. Feature importance for churn prediction: Random Forest-based ranking of behavioral predictors

Leveraging the continuous probability outputs, the system partitions the customer base into three actionable risk tiers[16], as illustrated in Figure 7:

• Low Risk (P_{churn} < 0.3, 49.7%): Subjected to baseline brand maintenance.

• Medium Risk (0.3 ≤ P_{churn} < 0.7, 46.0%): Triggers proactive, low-cost engagement.

• High Risk (P_{churn}≥ 0.7, 4.3%): Mandates immediate, high-incentive retention actions.

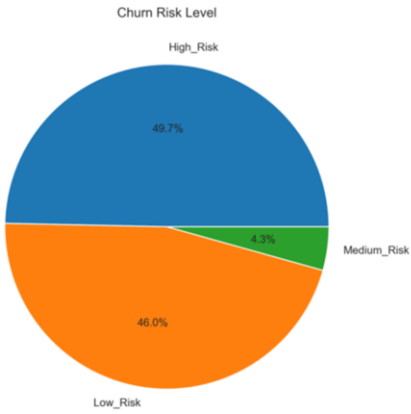


Figure 7. Churn risk level distribution: Three-tier customer risk classification

Churn Prediction Module Evaluation: To rigorously evaluate the predictive performance of the developed Random Forest model, the synthetic dataset was partitioned into an 80:20 split for training and historical validation. Unlike the baseline implementation which omitted quantitative metrics, this revised architecture evaluates the classification boundaries across five core dimensions: Accuracy, Precision, Recall, F1-Score, and the Area Under the Receiver Operating Characteristic Curve (AUC). Although accuracy is reported for completeness, model evaluation relies primarily on precision, F1, and AUC due to the inherent class imbalance in retail churn data. As tabulated in Table 2, the model delivers exceptionally robust performance.

The high Precision (0.919) is of paramount managerial significance; it implies a remarkably low false-positive rate, ensuring that expensive retention marketing resources are not squandered on customers who have no genuine intention of churning. Concurrently, the robust AUC (0.943) verifies the model's superior capability in ranking customer churn probabilities, establishing a reliable foundation for subsequent smart recall scheduling.

Table 2. Quantitative Performance Metrics of the Random Forest Churn Prediction Model.

Evaluation Metric	Value
Accuracy	0.880
Precision	0.919
Recall	0.850
F1-Score	0.883
Area Under Curve (AUC)	0.943

4.5. Personalized Recommendation Module

To optimize relevance and item diversity, the system deploys a three-layer hybrid recommendation topology.

Layer 1 - Segment-Driven Heuristics: Cold-start and structural constraints are handled by mapping items to user segments. Dormant cohorts are mapped to high-utility, cost-effective daily commodities; growing segments are served high-frequency trending products; while premium loyal tiers are matched with high-margin specialty items.

Layer 2 - User-Based Collaborative Filtering (UBCF): User similarity metrics are calculated over historical purchasing vectors using Cosine Similarity [17]:

$$sim(u,v) = \frac{\sum_{i\in I} r_{u,i} r_{v,i}}{\sqrt{\sum_{i\in I} r_{u,i}^2}\sqrt{\sum_{i\in I} r_{v,i}^2}} \tag{6}$$

where I represents the intersection of items rated by both users, and $r_{u,i}$ denotes the numerical interaction intensity (rating proxy) assigned by user u to item i. To generate recommendations, the unobserved preference score $\hat{p}_{u,j}$ for target user u on candidate item j is estimated via a weighted similarity aggregation across the top N local user neighborhood N(u):

$$\hat{p}_{u,j} = \bar{r}_u + \frac{\sum_{v\in N(u)} sim(u,v)\cdot(r_{v,j} - \bar{r}_v)}{\sum_{v\in N(u)} |sim(u,v)|} \tag{7}$$

where $\overline{r_u}$ and $\overline{r_v}$ characterize the baseline historical item consumption tendencies of users u and v, respectively. This uncovers latent preferences by cross-leveraging purchasing behaviors of peer groups.

Layer 3 - Market Basket Association Rules: The Apriori algorithm is executed over the transaction database to extract high-confidence itemsets, maximizing cross-selling efficiency [13]. Figure 8 visualizes the top 50 association rules by lift, revealing key product co-purchase patterns such as “Potato Chips + Coca-Cola” and “Tooth Paste + Tooth Brush.”

The outputs of these layers undergo weighted fusion and sorting. For example, low-value inactive cohorts are recommended high-turnover staples like Water, Potato, and Salt via segmentation-guided cost filters, whereas high-value loyal customers are presented with lifestyle cross-sell options like White Wine, Red Wine, and Cooking Oil.

Recommendation Module Empirical Validation: To thoroughly investigate the performance and algorithmic behavior of the personalized recommendation module, the transactional repository was subjected to a rigorous time-based evaluation split. Transactions prior to October 1, 2025 (comprising 1,955 purchase orders) were allocated to the training tensor to generate personalized candidates, while subsequent logs (comprising 1,410 purchase orders) were sequestered as the predictive ground truth. The evaluation measures accuracy and coverage across a strict top-K retrieval threshold (K = 3), tracking Precision, Recall, Hit Rate, Catalog Coverage, and Inter-User Diversity. The empirical outcomes are synthesized in Table 3.

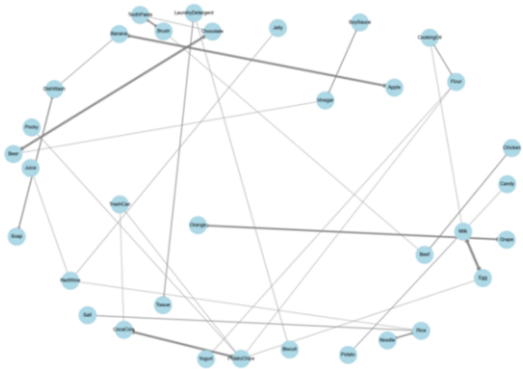


Figure 8. Product association network: Top 50 association rules by lift, illustrating cross-category purchase dependencies

The quantitative matrix reveals a highly nuanced behavior inherent to joint collaborative-heuristic recommendation spaces under strict structural constraints. At K = 3, the final Hybrid merged pipeline yields identical empirical metrics to the standalone Cluster-Based configuration. Crucially, this convergence does not imply algorithmic redundancy; rather, it uncovers a deterministic “heavy tail” saturation effect native to grocery retail environments. Within a constrained three-slot retrieval window, the segment-driven heuristics—which isolate high-volume, high-probability staple commodities—natively dominate the peak score ranking tiers due to their systemic baseline likelihood, effectively shielding the system against the extreme data sparsity that typically degrades pure collaborative pipelines. The high Inter-User Diversity (0.9011) does not arise from the segment-driven heuristics alone, but from reranking and micro-personalization introduced by collaborative filtering and association rules on top of a stable staple-item foundation. This mathematically guarantees that while the density of the top-3 retrieval is anchored on the most corporate-stable 18% of the SKU catalog to secure high

conversion, the micro-sorting and distinct item combinations across individual consumer profiles remain profoundly decentralized and personalized, delivering a robust defensive posture against catastrophic catalog concentration while maximizing capital return on short-window marketing inventories.

Table 3. Quantitative Evaluation and Ablation of the Recommendation Topologies (K = 3).

Recommendation Methodology	Precision@3	Recall@3	Hit Rate@3
Cluster-Based (Segment Heuristics)	0.2400	0.0392	0.4333
Collaborative Filtering (IBCF)	0.2278	0.0364	0.4033
Market Basket Association (Apriori)	0.2311	0.0400	0.4100
Hybrid (Merged Pipeline)	0.2400	0.0392	0.4333
Systemic Capacity Metrics (K = 3):			
		Catalog Coverage: 0.1800	Inter-User Diversity: 0.9011

4.6. Smart Recall Strategy Module

The strategy optimization engine uses segments and probabilities to dynamically trigger automated matrix responses:

• Priority Retention Tier (High-value, high-risk cohorts): Dispatches a 25% discount voucher coupled with exclusive SKU gifts and premium services.

• Standard Recall Tier (Medium-risk cohorts): Dispatches a 20% generic promotional voucher.

• Aggressive Win-Back Tier (Low-value, highly churned cohorts): Dispatches a high-incentive 35% deep discount to clear stagnant inventory.

The module automatically generates personalized copy leveraging these parameters, e.g., “Dear C0001, we miss you! Enjoy 35% OFF on Water, Potato, and Salt. Order now!”

This tiered discount design avoids blind marketing investment and effectively reduces unnecessary promotion cost compared with uniform full-customer discounts used in traditional CRM paradigms, establishing an economically robust financial baseline.

The asymmetric allocation of coupon discounts (e.g., 35% for low-value inactive cohorts versus 25% plus premium services for loyal cohorts) is rooted in marginal cost optimization. High-value loyal customers are sensitive to service quality and relationship maintenance, where excessive discounting would merely cannibalize profit margins. Conversely, low-value inactive customers represent stagnant assets with near-zero transactional contributions; deploying a high-stimulus 35% discount serves as a low marginal-cost reactivation test to probe their latent consumption willingness, maximizing overall return on marketing investment (ROI).

4.7 System Deployment Implementation

The functional prototype is engineered entirely within the Python ecosystem. Data processing and manipulation are handled via Pandas and NumPy; algorithmic clustering and ensemble modeling are powered by Scikit-learn; association rules are generated via mlxtend; and visualization layers utilize Matplotlib and Seaborn.

The production interface is built as a web application via Streamlit, structured into six interactive analytics modules: Store Overview, Customer Characteristic Analysis, K-Means Customer Segmentation, Churn Risk Analysis, Product Association Analysis, and Recommendation & Recall Strategy. An Automated Query Engine natural language interface is embedded to provide store managers with rapid conversational access to customer profiles and marketing metrics. This layout provides low-latency decision tools that eliminate manual SQL queries for front-line store coordinators.

5. Conclusions

This study has successfully developed and validated an integrated CRM decision-support architecture tailored for the supermarket retail sector by integrating a comprehensive data mining suite. The implemented architecture bridges data ingestion, multi-dimensional segmentation, predictive ensemble modeling, and hybrid item recommendation into a unified closed-loop infrastructure. By converting raw transaction history into interpretable behavioral insights, the framework offers an actionable roadmap for data-driven retail management.

With 41.7% inactive customers and 52.7% churn-prone users in the dataset, our integrated architecture provides targeted operational solutions to ease severe customer polarization pain points in supermarket retail.

Our empirical findings underscore a critical structural challenge in modern retail: a deep polarization within the customer base, where 41.7% of users are stagnant or inactive, and 52.7% exhibit active churn patterns. Crucially, feature importance ranking isolates purchase frequency as the primary operational bellwether for customer attrition (importance = 0.430), heavily outperforming demographic variables. Deploying this multi-tiered architecture onto an interactive Streamlit environment provides corporate managers with an interpretable, highly scalable decision-support instrument that effectively transitions enterprise operations from reactive auditing to proactive optimization.

6. Future Horizons

To further extend the capabilities and generalizability of the proposed intelligent CRM framework, several future research avenues are planned:

Empirical Field Validation: Integrating and testing the analytical system against high-volume, real-world historical membership datasets to evaluate algorithmic performance amid shifting market variables [18].

Temporal Feature Engineering: Developing sliding-window and rolling-trend attributes (e.g., 7-day, 30-day, and 90-day purchase volatility metrics) to more precisely capture real-time consumer behavior shifts.

Deep Learning Recommenders: Transitioning from traditional collaborative filtering to Neural Collaborative Filtering (NCF) and Transformer-based sequential recommendation architectures to better model temporal purchase trajectories.

LLM Knowledge-Base Integration: Replacing the basic rule-based QA engine with an agentic Large Language Model (LLM) framework capable of handling complex, multi-turn analytical queries and automatic report generation

Omni-Channel Analytics: Fusing in-store traffic mapping data into the feature repository to align physical actions with digital footprints.

References

[1] Kumar V, Reinartz W, Customer Relationship Management: Concept, Strategy, and Tools. Springer, 2018.

[2] Ngai EWT. Customer relationship management research (1992-2002): An academic literature review and classification. Marketing Intelligence & Planning. 2005;23(6):582–605.

[3] Linoff GS, Berry MJ. Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management. John Wiley & Sons, 2011.

[4] Gupta G, Aggarwal H. Improving customer relationship management using data mining techniques. International Journal of Machine Learning and Computing. 2012:2(6):874–877.

[5] Singh RA, Sharma SK. Data mining-driven customer relationship management: A systematic review. IEEE Access. 2024;12:56789–56807.

[6] Wang, RJC, Yen DC. Data mining techniques for customer relationship management. Technology in Society, 2002;24(4):483–502.

[7] Winer RS. A framework for customer relationship management, California Management Review. 2001:43(4):89–105.

[8] Ling SS, Too CW, Yen WWW, Chia LK. Customer Relationship Management System for Retail Stores Using Unsupervised Clustering Algorithms with RFM Modeling for Customer Segmentation. Proceedings of the IEEE International Symposium on Computer Applications and Industrial Electronics (ISCAIE); c2024. pp. 1–6, doi://10.1109/ISCAIE61308.2024.10576353

[9] Hassan RS, Nawaz A, Lashari MN, Aslam F. Effect of customer relationship management on customer satisfaction. Procedia Economics and Finance. 2015;23:563–567.

[10] Wei JT, Lin SY, Wu HH. Customer relationship management in the hairdressing industry: An application of data mining techniques. Expert Systems with Applications. 2013;40(18):7513–7518.

[11] Tsai CF, Chen KY. A hybrid approach for cluster analysis in customer relationship management. Expert Systems with Applications. 2010;37(7):4962–4966.

[12] Breiman L. Random forests. Machine Learning. 2001;45(1):5–32.

[13] Agrawal R, Srikant, R. Fast algorithms for mining association rules. Proceedings of the 20th International Conference on Very Large Data Bases (VLDB); c1994. pp. 487–499.

[14] Zheng Y. Customer relationship management and segmentation in retail grocery using machine learning and advanced RFM frameworks. IEEE Transactions on Engineering Management. 2023;70(4):1412–1423.

[15] Al-Mansoori, Al-Kabi, MN. A data mining framework for predictive churn analytics and personalized marketing in modern supermarket supply chains. Proceedings of the IEEE International Conference on Big Data (Big Data);2022. p. 2105–2114.

[16] Coussement K, Van den Poel D. Churn prediction in subscription services: An application of support vector machines and random forests. Expert Systems with Applications. 2008;34(1):245–253.

[17] Sarwar B, Karypis G, Konstan J, Riedl J. Item-based collaborative filtering recommendation algorithms. Proceedings of the 10th International Conference on World Wide Web;2001. p. 285–295.

[18] Soysal G, Zentner A, Zheng Z. Physical stores in the digital age: How store closures affect consumer churn. Production and Operations Management. 2019:28(11):2778–2781.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261109

Thinking from Quantum Computing Theories to Problem and Applications by Using A Novel Ontology-based Quantum Architecture Model

Yanyun LU[a] and Adela S.M. LAU[b,1]
[a] *Department of Commerce, Hong Kong Baptist University, China*
[b] *Data Science Lab, Department of Statistics and Actuarial Science, School of Computing and Data Science, The University of Hong Kong, China*
ORCiD ID:
Yanyun Lu https://orcid.org/0009-0006-8837-6855
Adela S.M. Lau https://orcid.org/0000-0001-5918-8309

Abstract. Quantum computing overcomes the limitations of computational power in classical computing by utilizing the principles of quantum mechanics; it has a transformative impact on improving computational effectiveness in data analysis. The existing industrial sectors are unclear how to apply theories and concepts into applications of quantum computing though there are many research studies. Therefore, this study aims to review the current state-of-the-art of quantum computing algorithms by discussing their mathematical frameworks, underlying architectures, and applications, and propose a novel ontology-based quantum architecture model for developers to select, filter and cascade different components for implementation. A case study is used to demonstrate the effectiveness of using the proposed model to design and implementation of a quantum computing algorithm with the proposed model. The results found the ontology model is easy to understand, ease of use, with available resources and technology facilitating for guiding developer on the development of a quantum application starting from problem definition to human problem-solving processes, and a quantum computing design.

Keywords. Quantum computing; data science; quantum applications; quantum case study.

1. Introduction

With the rapid development of classical computing in recent years, researchers have gradually realized that simply increasing computational power is insufficient for efficiently solving computational problems as their limited memory storage and long computation times. Quantum computing was introduced by Richard Feynman in 1982 and offers a promising solution to these computational bottlenecks [1]. By leveraging the principles of quantum mechanics, quantum computers exhibit inherent parallelism,

[1] Corresponding Author: Adela S.M. Lau, Data Science Lab, Department of Statistics and Actuarial Science, School of Computing and Data Science, Run Run Shaw Building, The University of Hong Kong, Pokfulam Road, Hong Kong; E-mail: adelalau@hku.hk.

enabling computational efficiency far beyond what classical computing systems can achieve. The resulting quantum advantage demonstrated the potential to revolutionize fields that classical computing finds intractable. However, the researchers mainly focus on the design of quantum computing algorithms and do not focus on the cognitive thinking of how the problems being able to be modeled into quantum computing. The companies and industrial sectors, especially in business and finance, do not understand how quantum computing can be used in business analysis, resulting in a slow pace and limited investment in the development of quantum applications.

Therefore, this paper aims to provide a fundamental introduction to the underlying principles of quantum algorithms. Several fundamental and well-known quantum algorithms and their applications are discussed. A case study of the implementation of a quantum algorithm is presented and discussed.

2. Literature Review

In 1982, Richard Feynman introduced the groundbreaking concept of a quantum computer, which simulated quantum systems and performed high-speed calculations [1]. Feynman's unconventional talk served as the intellectual spark that ignited the field of quantum computing. In the following decades, the quantum computing field experienced rapid growth, attracting brilliant minds from various disciplines, including physicists, mathematicians, computer scientists, and engineers. Their collective efforts have brought forth numerous groundbreaking theoretical and experimental advancements, pushing the boundaries of this interdisciplinary domain. Today, quantum computers have successfully entered some commercial uses and their immense potential and value continue to inspire researchers to further explore this revolutionary technology.

Quantum computing uses principles from quantum mechanics, like superposition and entanglement, to solve complex problems much faster than classical computers. Superposition is a quantum principle that a single particle can exist in multiple states simultaneously, until it is measured. Entanglement is a phenomenon where two or more quantum particles are linked and the state of one instantly influences the state of the other. In other words, by applying quantum theory, the superposition represents different concepts of an entity that depend on the context and situation (the state). The entanglement represents the relationships of the entities that are linked, in direct or indirect ways, to construct a scenario with a sequence of causes and consequences.

Quantum algorithms heavily rely on mathematical methods for their formulation and analysis. The most commonly used mathematical tools include linear algebra, probability theory, and complex analysis, along with more advanced concepts, such as Fourier transforms, game theory, computational complexity theory, and so on. The two fundamental elements of quantum algorithms are qubits and quantum gates [2][3]. Similar to classical computing, qubits are represented by using 0 and 1. However, the key difference lies in the fact that a classical bit can only be either 0 or 1 at any given time; whereas, a qubit, due to its state of superposition, can be both 0 and 1, simultaneously. The exact outcome depends on the measurement result. The notation for qubits represents 0 and 1 within the ket notation $|\cdot\rangle$, which can also be expressed as a vector, as shown in Equation (1).

$$|0\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, |1\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad (1)$$

Bra notation $\langle\cdot|$ represents the conjugate transpose of a ket, denoted by the dagger symbol $|\cdot\rangle^{\dagger}$, which is given by Equation (2):

$$\cdot| \;=\; |\cdot\rangle^{\dagger} \;\equiv\; (|\cdot\rangle^{*})^{T} \tag{2}$$

Using bra-ket notation to describe quantum states of qubits provides significant convenience in the derivation of complex quantum algorithms.

In different types of quantum computers, various physical quantities can be used to represent the qubit states $|0\rangle$ and $|1\rangle$. For example, in quantum dot quantum computers, the spin-up and spin-down states of an electron (a spin ½ particle) serve as qubit representations. In photonic quantum computers, the horizontal and vertical polarization of photons encodes the quantum information. In superconducting quantum computers, the presence or flow of electrons is used to define qubit states [4].

These two basis states $|0\rangle$ and $|1\rangle$ span the two-dimensional complex Hilbert space $\mathbb{C}^{2}$. For an n-qubit system, the corresponding state space is the tensor-product Hilbert space

$$\mathcal{H}_n = (\mathbb{C}^{2})^{\otimes n}$$

The computational basis of this space is given by:

$$\{\,|x\rangle\colon\; x \in \{0,1\}^{n}\}$$

where each basis vector $|x\rangle$ corresponds to a binary string of length n.

A pure quantum state of an n-qubit system is represented by a unit vector

$$|\psi\rangle \in \mathcal{H}_n, \qquad \langle\psi|\psi\rangle = 1$$

A quantum gate acting on an n-qubit system is represented by a unitary operator

$$U\colon \mathcal{H}_n \to \mathcal{H}_n, \qquad U^{\dagger}U = I$$

If the state before measurement is $|\psi\rangle$, then the probability of obtaining the basis state $|x\rangle$ in a computational-basis measurement is given by

$$p(x) \;=\; |\langle x \,|\, \psi\rangle|^{2}$$

The notation introduced in this subsection provides the mathematical foundation for the discussions of superposition, entanglement, quantum gates, quantum circuits, and the Quantum Fourier Transform in the following sections.

2.1. Superposition

Any single qubit in a quantum state $|\psi\rangle$ can be expressed as a linear combination of the computational basis states $|0\rangle$ and $|1\rangle$ in the Hilbert space H, as shown in Equation (3).

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle, \qquad \alpha, \beta \in \mathbb{C}, \tag{3}$$

Here, α and β are complex probability amplitudes that determine the likelihood of measuring the qubit in state $|0\rangle$ or $|1\rangle$, respectively, which are given by $|\alpha|^2$ and $|\beta|^2$. These probabilities satisfy the normalization condition $|\alpha|^2 + |\beta|^2 = 1$. A measurement in the computational basis yields the outcome $|0\rangle$ with probability $|\alpha|^2$ and $|1\rangle$ with probability $|\beta|^2$. Applying quantum theory into practice, if the market state is bullish ($|\alpha|^2=0$ and $|\beta|^2=1$), investors will change their purchasing intention to buy stock ($|\psi\rangle = |1\rangle$), and vice versa.

We can also represent any single-qubit state $|\psi\rangle$ in polar form, specifically on the Bloch Sphere, where $|0\rangle$ is the northern pole and $|1\rangle$ is the southern pole, as shown in Equation (4).

$$|\psi\rangle = \cos\frac{\theta}{2}|0\rangle \;+\; e^{i\emptyset}\sin\frac{\theta}{2}|1\rangle \tag{4}$$

We can gain a more intuitive understanding of the Bloch Sphere from Figure 1 [5]. The parameter θ controls the latitude (superposition) on the Bloch Sphere, determining the weighting of the basis states |0⟩ and |1⟩, while the parameter φ controls the longitude (global phase) on the Bloch Sphere, affecting the relative phase of the quantum state.

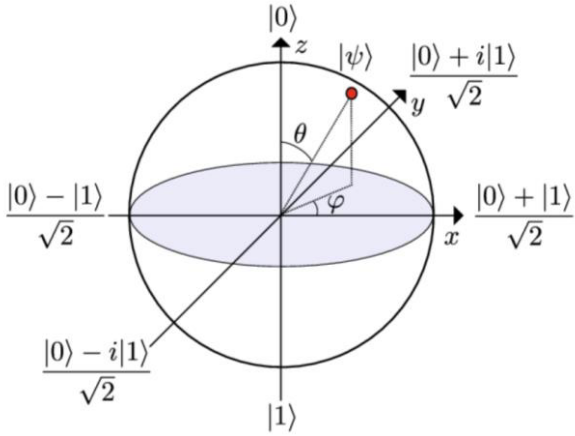


Figure 1. Bloch Sphere representation of a qubit.

Superposition is one of the fundamental principles that enable the computational efficiency of quantum algorithms. It grants quantum computers the ability to perform parallel computations, as qubits in a superposition encode information about multiple possible solutions simultaneously. Through quantum interference and final measurement, the superposition state collapses to yield the desired result.

2.2. Entanglement

Quantum entanglement is one of the most fundamental and intriguing phenomena in quantum mechanics. It describes a strong non-local correlation between multiple qubits, such that the quantum state of each qubit cannot be described independently but must be considered as part of a single entangled system. This property enables quantum algorithms and quantum communication protocols that surpass classical capabilities and further enhance quantum parallelism. Applying quantum theory into practice, entanglement describes how two or more entities correlate to construct a scenario environment for a new state. For example, a bullish stock market ($(|\psi_A\rangle \otimes |\psi_B\rangle)$) is constructed by the presence of good news ($|\psi\rangle_A$) and aggressive investors, who intend to buy stock ($|\psi\rangle_B$).

Mathematically, a quantum state can be written as the tensor product of two or more quantum states, such as $|\psi\rangle = |\psi\rangle_A \otimes |\psi\rangle_B$. The state $|\psi\rangle$ is a separable state, rather than an entangled state. The simplest form of entangled states for two qubits are the Bell states, which are given by Equation (5).

$$\left|\psi_A^{+}\right\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle), \qquad \left|\psi_A^{-}\right\rangle = \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle)$$
$$\left|\psi_B^{+}\right\rangle = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle), \qquad \left|\psi_B^{-}\right\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle) \tag{5}$$

Measuring one qubit of an entangled state instantaneously determines the state of the other, regardless of the spatial separation. The nature of quantum entanglement allows long-distance quantum communication.

2.3. Quantum Interference and Quantum Gates

The superposition states of qubits describe the inherent uncertainty in quantum mechanics. Quantum interference provides a mechanism to enhance the amplitude of the desired output, whilst suppressing the amplitude of incorrect outputs through the application of quantum gates.

The quantum gates [2] are used to manipulate qubits in a quantum computer. In other words, applying quantum theory into practice, quantum interference acts as a set of mathematical operators to compute a quantum state, while quantum gates represent the mathematical equation in a matrix notation. A stock forecasting regression model applies the summation operators (quantum interference) to the market factors to compute the buy or sell signal (the 1 or 0 state). The mathematical equation (quantum gate) presents the relationships of the factors of a specific market scenario for the stock forecasting model.

In linear algebra, a quantum gate is expressed as a 2^n x 2^n matrix, where n is the number of qubits it operates on. Table 1 includes some common quantum gates, their symbols, matrix representations, and operations. Pauli-X Gate acts like a classical NOT gate. Pauli-Z gate provides a phase flip to the qubit without changing its state. Pauli-Y gate introduces a phase shift and flips the qubit state, and is a combination of the X and Z gates. The Hadamard gate is designed for quantum parallelism. It creates superposition that a qubit can exist in both $|0\rangle$ and $|1\rangle$ simultaneously and with equal probability. It is one of the most important quantum gates, enabling the transformation of $|0\rangle$ into the plus state, as shown in Equation (6) and (7). The phase gate adjusts the relative phase of a qubit's quantum state by rotating the state vector around the Z-axis of the Bloch sphere without changing the probabilities of measuring its 0 or 1 values, and is commonly used for solving complex problems. The $\pi/8$ gate is for adding the phases to allow the gate to reach any valid quantum state. The CNOT gate has two qubits in the gate. The first qubit is for the control the flipping of the state of the second qubit. The controlled-z gate is for. The SWAP gate rearranges the qubits and exchanges the states of them.

Table 1. Common quantum gates with their symbols, matrix representations and operations on qubits.

Quantum Gates	Symbol	Matrix	Operation
Pauli-X (X)		$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$	$X\|0\rangle = \|1\rangle$ $X\|1\rangle = \|0\rangle$
Pauli-Y (Y)		$\begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}$	$Y\|0\rangle = i\|1\rangle$ $Y\|1\rangle = -i\|0\rangle$
Pauli-Z (Z)		$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$	$Z\|0\rangle = \|0\rangle$ $Z\|1\rangle = -\|1\rangle$
Hadamard (H)		$\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$	$H\|0\rangle = \frac{1}{\sqrt{2}}(\|0\rangle + \|1\rangle) = \|+\rangle$; $H\|1\rangle = \frac{1}{\sqrt{2}}(\|0\rangle - \|1\rangle) = \|-\rangle$
Phase (S)		$\begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix}$	$S\|0\rangle = \|0\rangle$ $S\|1\rangle = i\|1\rangle$
$\pi/8$ (T)		$\begin{bmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{bmatrix}$	$T\|0\rangle = \|0\rangle$ $T\|1\rangle = e^{i\pi/4}\|1\rangle$
CNOT (CX)		$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$	$CX_{0,1}$: If qubit 0 is 0, do nothing; if qubit 0 is 1, qubit 1 flips.
Controlled-Z (CZ)		$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$	$CZ_{0,1}\|11\rangle = -\|11\rangle$; otherwise, do nothing.
SWAP		$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$	SWAP$\|ab\rangle = \|ba\rangle$

$$|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle \quad (6)$$

and $|0\rangle$ into the minus state, as shown in Equation (7).

$$|-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle \quad (7)$$

These two states are the most typical superposition states in quantum computing.

2.4. Quantum Circuit

A quantum circuit is a model for quantum computation [6], where qubits are manipulated using a sequence of quantum gates. The representation of a quantum circuit can be expressed using both symbolic diagrams and mathematical equations. Figure 2 illustrates a quantum circuit that transforms the initial state $|00\rangle$ into the Bell state $|\psi^+\rangle$. This process involves three quantum gates:

- The Hadamard gate (H) acts on the first qubit, creating a superposition state.
- The Pauli-X gate (X) flips the second qubit from $|0\rangle$ to $|1\rangle$.
- The Controlled-Not ($CX_{0,1}$) gate then entangles the two qubits, which can lead to the formation of the Bell state on the right-hand side of the circuit.

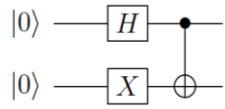


Figure 2. Quantum circuit and mathematical derivation for generating the Bell state $|\psi^+\rangle$.

Equation (8) is the expression of the quantum circuit.

$$CX_{0,1}(H \otimes \mathrm{X})(|0\rangle \otimes |0\rangle) = CX_{0,1}\frac{1}{\sqrt{2}}(|01\rangle + |11\rangle) = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle) = |\psi^+\rangle \quad (8)$$

By applying quantum theory into practice, the quantum circuit formulates the consequence of a complex scenario. In the case of stock market prediction, there is no single category of market factor (e.g., technical analysis, fundamental analysis, or fear index analysis) to predict the stock price. The market problems are broken down into sub-problems, such as a fundamental problem, technical problem, and investor behavior problem. The cause and consequence of each problem can be formulated in a quantum gate or a set of quantum gates. In other words, the prediction model (quantum circuit) is constructed using a set of quantum gates with a hierarchical structure, to present the logical thinking process of problem solving.

2.5. Theory of Quantum Algorithms and Their Applications

In 1985, David Deutsch, known as the father of quantum computing, first proposed the theoretical model of the Quantum Turing Machine in his paper [7]. In this work, the Church-Turing-Deutsch thesis states:

"Every finitely realizable physical system can be perfectly simulated by a universal model computing machine operating by finite means."

The feasibility of a universal quantum computer was thus established. Classical computers are fundamentally limited in their ability to simulate all physical processes, especially quantum systems, due to the exponential growth in computational and storage resources required. However, using quantum mechanics itself to simulate quantum systems, namely quantum computing, is a more reasonable approach, which can ensure computational completeness in the physical sense.

David Deutsch's paper laid the foundation for the subsequent development of quantum algorithms, including the famous Grover's Algorithm [8] and Shor's Algorithm

[9]. This work also established the fundamental principle that the oracle operator U_f must be reversible in quantum algorithms, to comply with the unitary nature of quantum mechanics. The fundamental architecture of this oracle are described in Equation (9) and Figure 3.

$$U_f(|\mathrm{x}\rangle_n|\mathrm{y}\rangle_m) = |\mathrm{x}\rangle_n|y \otimes \mathrm{f(x)}\rangle_m \tag{9}$$

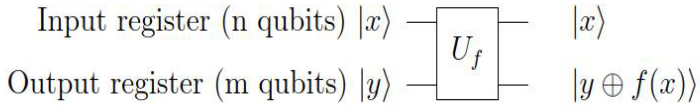


Figure 3. Quantum circuit representation of the oracle U_f along with its corresponding mathematical formulation.

Here, the function f(x) is a classically computable function, which is typically well-defined. The oracle U_f consists of specific quantum gates tailored for different f(x), i.e., different computational purposes, to enable the efficient evaluation of f(x). Therefore, quantum algorithms designed within this framework are often specialized for particular problem domains. The Deutsch-Jozsa Algorithm [10] was the first quantum algorithm for searching problems demonstrating the advantage of using quantum over the classical computing of $O(2^n)$. Researchers have also applied different quantum algorithms for cryptanalysis, eigenvalues computation, data analysis, and financial predictions. A group of researchers used quantum Fourier transform for cryptanalysis [11], while Shor's Algorithm was used to break the RSA encryption method [12]. Another group of researchers used Quantum Phase Estimate [13] for eigenvalues in quantum chemistry and to solve linear system problems [14][15]. Variational Quantum Eigensolver [16] has been used in quantum chemistry and material science [17][18]. Some researchers developed the Quantum Support Vector Machine model [19] for classification [20][21]and used Quantum Monte Carlo [22] for computing the probability of financial risk [23][24]. However, they did not discuss the thinking processes of developing a quantum computing circuit starting from problem itself. It makes developer not able to design a more complicated quantum computing circuit for the complicated problems. A more systematic way of explanation and guidance is required for developers, especially for those non-technical personals, to understand the basic components of the quantum computing, and how they being connected to build the circuit for problem solving. Table 2 summarizes the problem definitions and quantum algorithms for these applications.

Table 2. Overview of the applications of quantum algorithms.

Algorithm	Problem Definition	Quantum Complexity	Output	Explanation	Application
Quantum Fourier Transform [25]	Perform the discrete Fourier transform on a quantum state of size $N = 2^n$.	$O(n^2)$ gates.	Fourier transformed state.	Foundation for many quantum algorithms (Shor's, QPE).	Period finding, phase estimation, cryptanalysis [11 .
Shor's Algorithm[26]	Given an integer M, find its prime factors.	Polynomial in log M (e.g., $((\log M)^3)$).	Prime factors of M.	Use QFT to find the period of a function mod M.	Breaking RSA encryption, cryptanalysis

					[12].
Quantum Phase Estimation (QPE) [13]	Estimate the eigenvalue (phase) of a given unitary U.	Polynomial in problem size and 1/ε.	Approximate eigenphase of U.	Use QFT to measure the phase accurately.	Finding eigenvalues in quantum chemistry, solving linear systems [14,15].
Variational Quantum Eigensolver (VQE)[16]	Approximate the ground state energy of a Hamiltonian.	Hybrid approach feasible on near-term devices.	Approximate ground-state energy.	Combine a parameterized quantum circuit with a classical optimizer.	Quantum chemistry, materials science [17,18].
Quantum Support Vector Machine [19]	Classify data using quantum kernels or feature maps.	Possible exponential speedups with suitable encoding.	Classification result.	Encode data in quantum states for kernel-based learning.	Machine learning, data analysis [20,21].
Quantum Monte Carlo [22]	Estimate integrals or expected values with Monte Carlo sampling.	O(1/ε) using amplitude estimation (quadratic speedup).	Approximate integral or expectation value.	Use amplitude estimation to achieve faster convergence than classical methods.	Finance (risk analysis), physics simulations [23,24].

3. A Novel Ontology-based Quantum Architecture Model for Quantum Algorithm Development

Therefore, a novel ontology-based quantum architecture model was proposed and was shown in Figure 4.

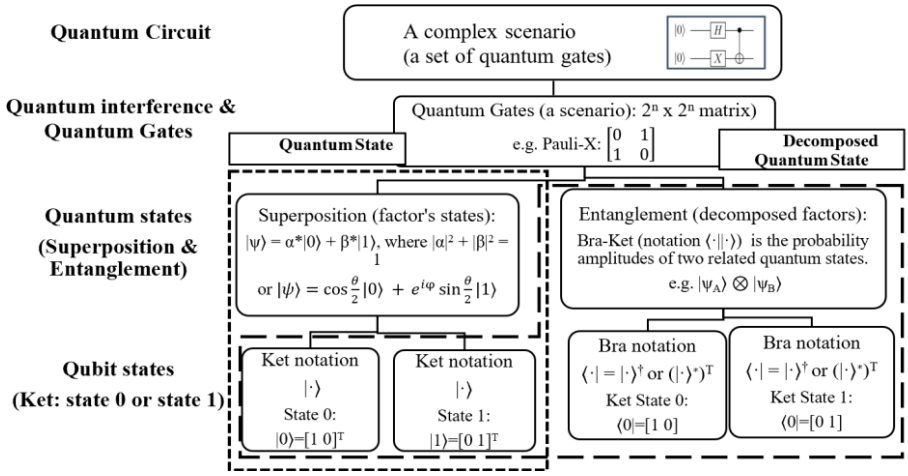


Figure 4. An ontology-based quantum architecture model.

In the model, a quantum circuit contains a set of quantum gates and is used to represent the complex scenario, which can break down into several problem scenarios, in the real world. The quantum gate represents one problem scenario which contains the problem context objects, problem solving steps, and the problem objects' status

caused by multiple factors. The quantum interference breaks down the complex scenario into different scenario problems [27] and select different quantum gates to represent the scenario problems. The quantum gate's status depends on many factors in the problem scenario to make its status change. To break down the problem scenario into several problem-solving steps, the entanglement process models how different problem objects influenced by multiple factors, models the problem-solving steps and the cause-and-consequence logics, and models these logics into the mathematical equations (i.e. objects' convolution). The superposition of the quantum status describes how the quantum status changes with the problem factors status's change (i.e. the change of ɵ and φ), and uses the mathematical equations to model the quantum state's changes. The quantum state in the gate is represented by a qubit, which has the states of 0 or 1 to describe the outcomes of the event.

In other words, the ontology mapping across the problem-solving steps and quantum computing architecture design can be described in figure 5.

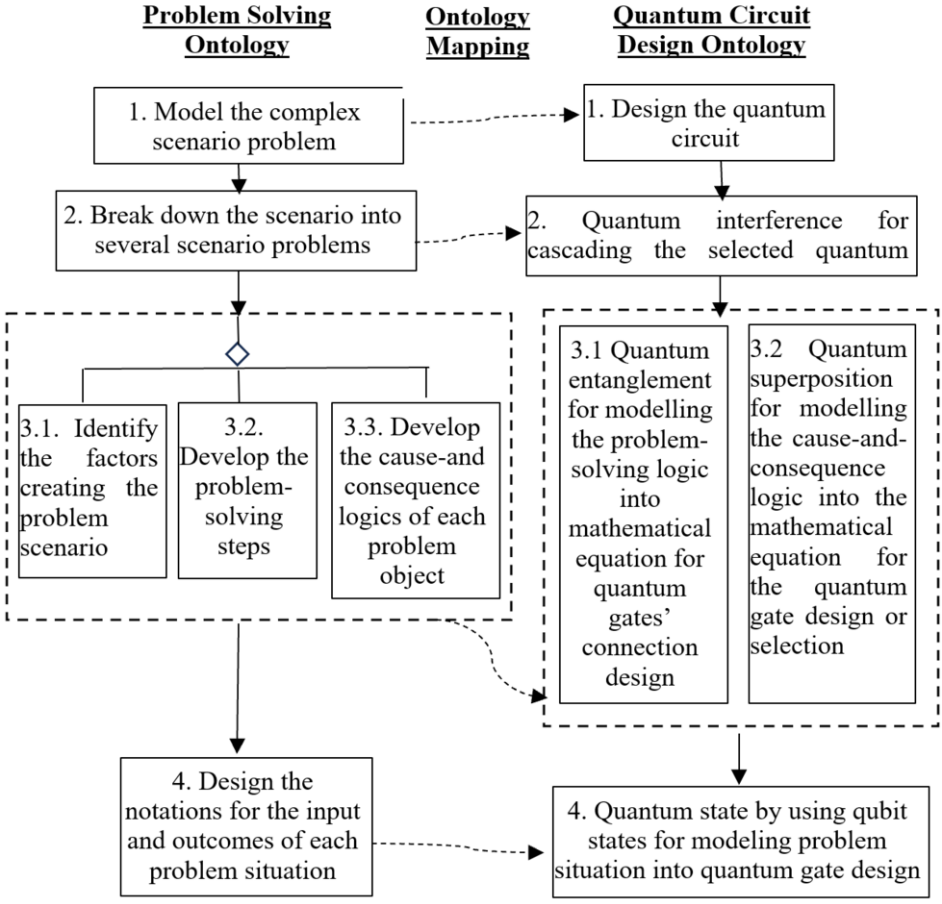


Figure 5. Ontology mapping of the problem-solving process and the quantum algorithm design.

4. Research Method

In order to demonstrate how to use the ontology model to design a quantum algorithm for an application, developer adopted the ontology-based quantum architecture model to model its thinking process and apply it to the design of the simulation and measurement of a quantum algorithm using Python as a case study. We wrote a program, as an example of Quantum Fourier Transform, to change an input signal from one state to another state by changing its time frequency. The simulation experiment was implemented with the Qiskit Python library [28], which is a powerful and convenient tool for quantum computing simulations and operations, and accessible to real quantum computers through the IBM Quantum API.

We ran the simulation program on a local computer and in real quantum computing hardware from the IBM Quantum Cloud Platform, which provides quantum processors and high-performance cloud-based execution services, in order to evaluate execution performance in both environments. The lessons learned from its implementation and the thinking processes of applying Quantum Fourier Transform algorithm to a stock forecasting problem are discussed below.

5. The Case Study on Quantum Fourier Transform Implementation

We used the Qiskit Aer simulator and real IBM quantum hardware to implement the Quantum Fourier Transform (QFT). The QFT is one of the core algorithms in quantum computing and it is the quantum version of the classical Fourier Transform (FFT). QFT transforms the input quantum state into its frequency domain representation. Its primary characteristic is achieving exponential speedup through quantum interference. The mathematical representation of the QFT for a single qubit state x in an n-qubit system is given by Equation (10).

$$QFT(|\mathrm{x}\rangle) = \frac{1}{\sqrt{2^n}} \sum_{y=0}^{2^n-1} e^{\frac{2\pi i}{2^n} xy} |\mathrm{y}\rangle \qquad (10)$$

We constructed a 3-qubit quantum circuit to implement the QFT on the input state $|000\rangle = |0\rangle$, transforming it into a QFT representation of $|111\rangle = |7\rangle$, which is $|\tilde{7}\rangle$, as shown in Figure 6.

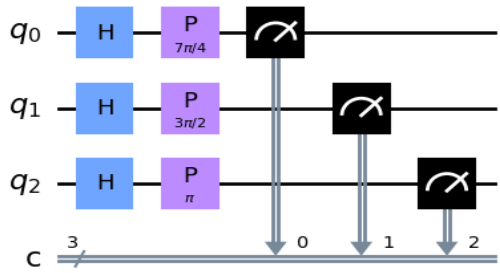


Figure 6. Quantum circuit for QFT.

The QFT transforms a given quantum state into its corresponding QFT representation. In our example, the transformation is from $|111\rangle$ to QFT($|111\rangle$). Therefore,

we first prepare the state |111⟩ from the initial state |000⟩ by applying an X gate to each qubit.

The quantum circuit is redrawn in black-and-white to ensure consistency with the existing figure style of the paper. In Qiskit version 2.3.1, built-in functions for QFT and inverse QFT are available. The corresponding QFT circuit is shown in Figure 7.

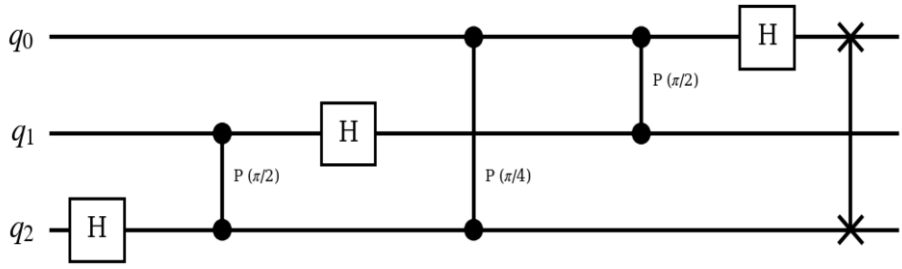


Figure 7. The QFT circuit.

For the measurement results obtained from both the simulator and the quantum computer, the sample code for the quantum circuit designed to implement the QFT is given in Figure 8. The output histogram is shown in Figure 9.

```
# *** Import all libraries ans set parameters ***
# ------------------------------------------------------------------------------------------
import matplotlib.pyplot as plt
from qiskit import QuantumCircuit, transpile
from qiskit.circuit.library import QFTGate
from qiskit_aer import AerSimulator
from qiskit.visualization import plot_histogram
from qiskit.transpiler import generate_preset_pass_manager
from qiskit_ibm_runtime import QiskitRuntimeService, SamplerV2 as Sampler

# parameters
nqubits = 3
shots = 2048
IBM_TOKEN = "YOUR_IBM_QUANTUM_API_TOKEN"
IBM_INSTANCE = "YOUR_IBM_QUANTUM_INSTANCE"

# *** Set up QFT using Qiskit's built-in function ***
# ------------------------------------------------------------------------------------------
def qft(circuit, n):
    circuit.append(QFTGate(n), circuit.qubits[:n])
    return circuit

# *** Build circuit: |111> -> QFT(|111>) ***
# ------------------------------------------------------------------------------------------
qc = QuantumCircuit(nqubits, nqubits)
```

```
    # Prepare |111>
    for i in range(nqubits):
        qc.x(i)

    # Apply QFT
    qft(qc, nqubits)

    # Add measurements
    qc.measure(range(nqubits), range(nqubits))

    # *** Run the local simulator ***
    # ---------------------------------------------------------------------------------------
    sim = AerSimulator()
    tqc_sim = transpile(qc, sim)
    result_sim = sim.run(tqc_sim, shots=shots).result()
    counts_sim = result_sim.get_counts()
    print("Simulator counts:")
    print(counts_sim)

    # *** Connect to IBM Quantum service ***
    # ---------------------------------------------------------------------------------------
    service = QiskitRuntimeService(channel="ibm_quantum_platform",token=IBM_TOKEN,instance=IBM_INSTANCE)
    backend = service.least_busy(simulator=False,operational=True,min_num_qubits=nqubits)

    # *** Transpile the circuit for the selected quantum backend ***
    # ---------------------------------------------------------------------------------------
    pm = generate_preset_pass_manager(optimization_level=1,backend=backend,translation_method="translator")
    isa_qc = pm.run(qc)

    # *** Run the circuit on real IBM quantum hardware ***
    # ---------------------------------------------------------------------------------------
    sampler = Sampler(mode=backend)
    job = sampler.run([isa_qc], shots=shots)
    result_qpu = job.result()
    pub_result = result_qpu[0]

    # Extract counts from the classical register
    creg_name = isa_qc.cregs[0].name
    counts_qpu = getattr(pub_result.data, creg_name).get_counts()
    print("Quantum computer counts:")
    print(counts_qpu)
```

Figure 8. Sample code for the quantum circuit designed to implement the QFT.

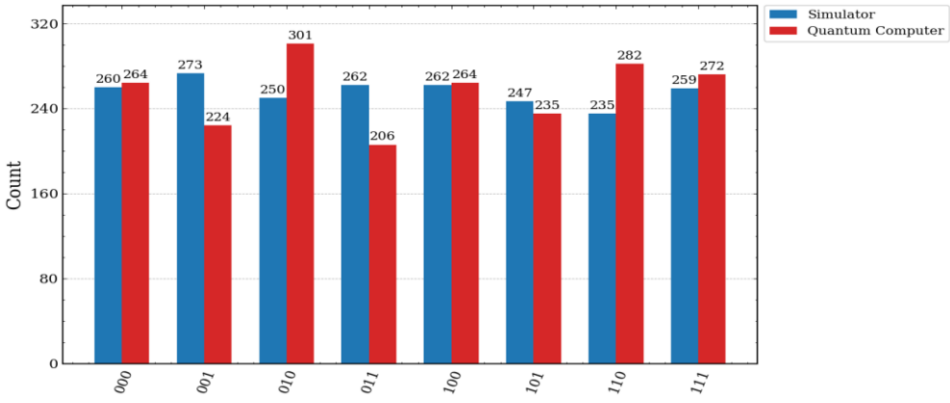


Figure 9. The output histogram.

For the simulator, the measurement is reasonable, since $|\tilde{7}\rangle$ has equal amplitude over all computational basis states in a 3-qubit system but only global phase differences, which can be expressed as in Equation (11).

$$|\tilde{7}\rangle = \frac{1}{\sqrt{8}}\sum_{k=0}^{7} e^{2\pi i\frac{7k}{8}|k\rangle} = \frac{1}{\sqrt{8}}[1|0\rangle + \frac{\sqrt{2}}{2}(1-i)|1\rangle - \mathrm{i}|2\rangle - \frac{\sqrt{2}}{2}(1+i)|3\rangle - 1|4\rangle + \frac{\sqrt{2}}{2}(-1+i)|5\rangle + i|6\rangle + \frac{\sqrt{2}}{2}(1+i)|7\rangle] \quad (11)$$

However, a quantum computer produces a different result when comparing with the simulator. This result may be attributed to error correction and quantum noise in real quantum computing systems.

The Inverse Quantum Fourier Transform (IQFT) is the inverse operation of the QFT. It transforms quantum states from the frequency domain back to the computational basis states and is commonly used to recover the original information from the transformed states. We apply the inverse QFT. The corresponding quantum circuit is shown in Figure 10.

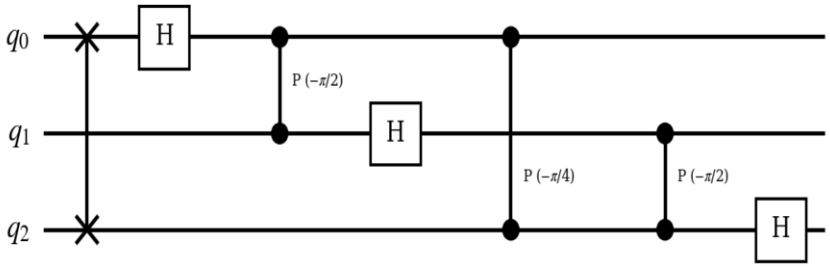


Figure 10. The quantum circuit of the inverse QFT.

The output histogram of the inverse QFT is shown in Figure 11. It seems that IBM has improved its quantum hardware, and the results obtained from the quantum computer are closer to the ideal simulations.

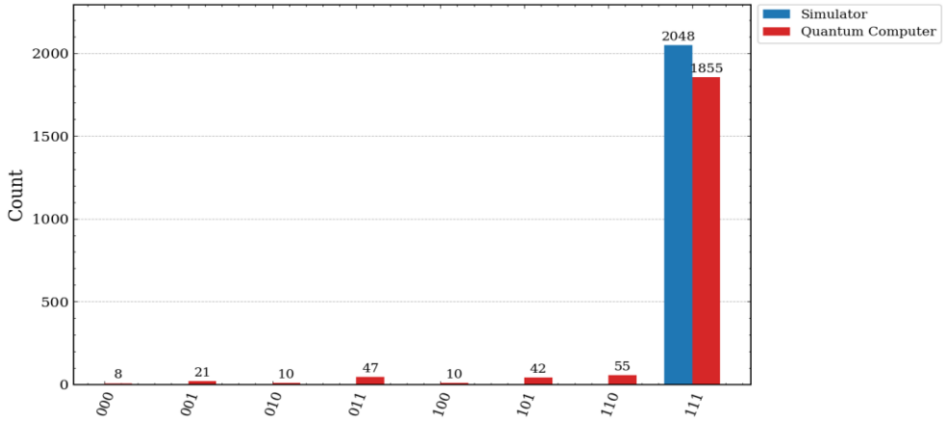


Figure 11. The output histogram of the inverse-QFT.

6. Discussion and Evaluation

From the results, we observe that all simulation results are measured as $|111\rangle = |7\rangle$. Meanwhile, the majority of results from the quantum computer are also measured as $|7\rangle$, while a small number of results correspond to other states. Therefore, the quantum simulator can be used to simulate quantum computing hardware to reduce the development cost. However, real quantum hardware is affected by quantum errors. Environmental noise, including thermal noise, electromagnetic radiation, and mechanical vibrations, along with decoherence and measurement errors, constitutes the primary source of quantum errors [29][30]. Therefore, when designing a quantum computing circuit, these sources of noise should be considered, in order to fine tune the simulated model; to date, no researchers have added environment error adjustment into the quantum computing algorithm and the quantum circuit design, to fine tune the measurement results from a simulated environment into a real environment. Therefore, the simulator can only be used in the initial design of a quantum computing circuit, to overcome the limited quantum computing resources [31]. Reducing quantum errors and the use of a quantum computing algorithm, to fine tune the model and improve the simulation, are some future research topics for scientists. In fact, machine learning experts with experience in training the parameters of the AI model. After they understand the mechanisms and limitations of quantum computing, they can apply their error adjustment methods and experience to the development of quantum algorithms.

To apply quantum computing in financial markets, Fourier transform can be used for stock price forecasting [32][33]. It can be used to determine the price fluctuation cycle [31] and to select the influencing factors for building the stock price prediction model [32]. In financial markets, the stock price goes up and down, depending on many factors. Each factor can be explained by some decomposing factors and has a different intensity of influence on the price fluctuation. The combination of these factors constructs the complex market scenario. Traditional methods for predicting stock price

include the stochastic model [34] , regression model [35], neural network model [36][37] , time series model [38] (Xu & Zhang, 2021) and Fourier transform [32][33].

The stochastic model makes predictions about outcomes using random factors. Each probability and random process is uniquely associated with a random variable in the market. A probability distribution is used to simulate the market scenario for stock price prediction. However, the stock market is influenced by the behavior of investors. Changes in the economic and political environment, together with the noise from market news, influence price change, which leads some researchers to believe that stock price is predictable. Therefore, some researchers use a regression model to identify the key factors for improving the accuracy of stock price prediction. However, the market factors are not linearly correlated and the price has a periodical fluctuation cycle due to investors' investment behavior and the investment cycle. Some other researchers have used a neural network to learn the non-linear price patterns for building a stock price prediction model. However, it consumed a lot of computational resources in training the model and could not identify the market's periodical cycles within the model. Although time series can be used to identify the market's periodical cycle, it cannot filter the noise and update the market changes instantly. Fourier transform is one of the techniques that has recently been used to identify noise and determine the periodical patterns [33]; it can be used to filter market noise and identify the major factors for building the predictive model to improve prediction accuracy [32].

However, the market environment changes very fast and investors require instant responses to market changes. Traditional computing has limited computational power and time resources to compute and update such changes. With the use of quantum Fourier Transform, the computation time can be shortened to make real time responses that react to market changes. By applying quantum computing to stock price prediction, the quant can represent a market factor state, which can be a positive state ("1") or a negative state ("0"), depending on the market conditions (the superposition). The quant can be broken down into two decomposed factors and notated by a ket and a bar. The ket is associated with its decomposed factors (the bar), where the bar and ket have the same common attributes and construct the market factor state. As an example, if investors' behavior in the market is aggressive buying (qubit state: 1), then this phenomenon can be explained by the fact that the current investors in the market (the ket) are risk-takers. If the current investors in the market like to buy risky investment product, and they all would like to put one million dollars to invest in investment products, and the investment products in the market (the bar) is risky stocks and they require a minimum entrance fee of 1 million dollars, then the dot product of ket and bar (the entanglement, where the bar is the transpose of the ket) models the single value decomposition relationship of the investors' buying behavior. In other words, the aggressive buying behavior of investors can be explained by the presence of the risky and wealth investors, who buy the risky and high entry fee stocks to make the market aggressive, with respect to buying.

A market state (e.g., market prices increasing or decreasing) can be explained by a set of market scenarios (quantum gates), including influencing macro-economic factors [39], influencing micro-economic factors [40][41], and investors' behavior [42]. The level of the influence (quantum interference) on the price changes depends on how investors react to the changes of those factors towards the market price, based on past experience. Quantum interference acts as an operator, linking and aggregating all the factors together, to build a prediction model. This prediction model (quantum circuit) is finally constructed by the mathematical algorithm that models the influences of the market factors (quantum gates). In other words, the quantum gates are the base

components of the quantum computing circuit that models the logical thinking, reasoning, and deducing of the impacts of those market factors towards the price changes; this constructs the market scenario equations for prediction and decision making.

In the previous example of quantum computing programming (see Figure 6), Fourier transform successfully transforms the inputs' states from 000 to 111 with three quantum gates, where each gate represents a quant state (a ket). The first gate represents the first digit of "000". Applying quantum computing into practice, the first, second and third gate (the bits) can represent the fundamental analysis, technical analysis, and fear index analysis. The results of the analyses change the quit state from 0 to 1, to reflect the market atmosphere changing from negative to positive (i.e., from a selling signal to a buying signal). This change can be explained by its decomposed factors. For example, the fundamental analysis (the quant state) depends on the P/E ratio, sales volume, cash flow, and debt values for each company. After further decomposition of the fundamental analysis, regarding these common attributes, two decomposing factors (the fundamental analysis decomposing factor (the ket) and the company decomposing factor (the bar)), are formed. Values are computed for the P/E ratio, sales volume, cash flow, and debt in the fundamental analysis, to form the "fundamental analysis decomposing factor vector" (the ket notation matrix). We extracted the company annual report data (the bar) for the values of P/E ratio, sales volume, cash flow, and debt, to form the "company decomposing factor vector" (the bar notation matrix). By performing the dot product of both vectors, the fundamental analysis vector was computed to reflect the fundamental analysis factor state (the quant state). In other words, the dot product of the two decomposing factors explains how they influence the fundamental factors' state for explaining a bull or bear market phenomenon. Similarly, the technical analysis can be further decomposed into the two decomposing factors, which are company stock and Hang Seng index; both of which have the common attributes of RSI, moving average, and beta, to explain the technical analysis state. The fear index analysis can be decomposed into two decomposing factors: risky investors and risk-averse investors. These factors have the common attributes of risk taker index, fear and greed index and news sentiments, which explain the fear index analysis state.

In Figure 6, we used three Hadamard gates and controlled phase rotation gates to construct the quantum circuit to represent the Quantum Fourier Transform algorithm. The Hadamard gates change the quant state by using the superposition mechanism and the three controlled phase rotation gates are applied in a specific order between qubits, to encode the frequency components required to change the quant state. Finally, swap gates were used to arrange the qubits in the correct order for the quant output state, where necessary. By using the Quasit Aer simulator with python, we constructed the quantum circuit to simulate the negative market environment with three quantum gates at the states of "000", which implies that the results of the fundamental analysis, technical analysis and fear index analysis are negative. We applied the phase rotation and superposition mechanism on the gates to change all of the values of the quant states from 0 to 1, to reflect the market state changing from negative ("000") to positive ("111"). In other words, the phase rotation and superposition mechanism in the quantum circuit models the logics of the market change, regarding the fundamental, technical and investor behavior conditions. To further model a more complicated market situation, we could use several quantum mathematical models to construct the quantum circuit and simulate the more complex market scenarios.

Therefore, researchers and software developers require a deep understanding of the mathematical models, quantum computing algorithms, quantum computing base models

(and their architectural framework), and the logical operations of quantum computing for developing quantum computing applications. Since we have limited quantum computing base components, more quantum mathematical models are required from researchers, to further develop and build a scalable and powerful quantum computing architecture and enable developers to design quantum circuits, especially for quantum artificial intelligence. Since artificial intelligence has grown rapidly over the last century, applying quantum design concepts to model the artificial intelligence model is urgently needed for the growth of the quantum computing industry.

Other than evaluating the effectiveness of the ontology model to guide the developer to design the quantum computing circuit with the case study, the theory of planned behavior model [43] was used to assess the concerns of users on the adoption of the ontology model to develop the quantum computing circuit. The constructs of the user concerns on the adoption includes the attitude and perceived behavioral control. The attitude includes the sub-constructs of perceived usefulness, relative advantage, easy to understand, and perceived ease of use. The perceived behavioral control includes the sub-constructs of resource facilitating and technology facilitating conditions. After evaluation from the developer on the adoption of the ontology, the preliminary result found that the ontology model is easy to understand and perceived ease of use to guide the thinking process of developing the quantum computing circuit. The resource facilitating and technology facilitating conditions of mapping the problem-solving steps into the quantum gates are satisfied as the basis unit of the quantum gate has already well implemented in various quantum computing libraries by software developers.

7. Conclusions

In conclusion, this paper introduced the fundamental mathematics and architecture of quantum algorithms and discussed examples of specific quantum algorithms, including their target problem definitions, architectural design, mathematical principles, and applications. It discussed the thinking process of how to develop a quantum computing application from the problem components, to the quantum computing circuit design, and to the application. We presented a case study which implemented a quantum algorithm and demonstrated how the mathematical model can be implemented into quantum computing. Suggested future research could be to provide more other application examples, such as simulating the mechanism for discovering material superconductors, combining different molecules to design more effective medicines, unravelling the mysteries of DNA to better understand ourselves, and training more powerful quantum machine learning models to assist different fields of applications with quantum computing. The leap in computational power brought by quantum computing has the potential to create groundbreaking changes in science, technology, and human life. Other than that, more experiments to evaluate the effectiveness of the ontology model are required to assess whether the ontology can guide developers to implement the quantum computing circuit with a better accuracy, execution time, circuit performance, and etc.

References

[1] Feynman RP. Simulating physics with computers. International Journal of Theoretical Physics. 1982;21(6–7):467-488, doi.org/10.1007/bf02650179.

[2] Mermin ND. Quantum Computer Science: An Introduction. Cambridge: Cambridge University Press. 2007; https://www.cambridge.org/core/books/quantum-computer-science/66462590D10C8010017CF1D7C45708D7.
[3] Nielsen MA, Chuang, I., Grover, L.K. Quantum Computation and Quantum Information. American Journal of Physics. 2022;70(5),558-559, doi.org/10.1119/1.1463744.
[4] Nandhini S, Harpreet S, Akash UN An extensive review on quantum computers. Advances in Engineering Software. 2022;174:103337, doi.org/10.1016/j.advengsoft.2022.103337.
[5] Nori K, Kockum AF, Nori F. Quantum Bits with Josephson Junctions. Fundamentals and Frontiers of the Josephson Effect. 2019;703–741, doi.org/10.1007/978-3-030-20726-7_17.
[6] Krantz P, Kjaergaard M, Yan F, Orlando TP, Gustavsson, S., Oliver, W.D. A quantum engineer's guide to superconducting qubits. Applied Physics Reviews. 2019;6(2):021318, https://doi.org/10.1063/1.5089550.
[7] Deutsch D. Quantum theory, the Church-Turing principle and the universal quantum computer. Proceedings of the Royal Society of London A. 1985;400:97-117, https://www.daviddeutsch.org.uk/wp-content/deutsch85.pdf.
[8] Grover LK. A fast quantum mechanical algorithm for database search. 28th Annual ACM Symposium on the Theory of Computing. 1996;212-219, https://arxiv.org/abs/quant-ph/9605043.
[9] Shor PW. Algorithms for quantum computation: discrete logarithms and factoring. Proceedings 35th Annual Symposium on Foundations of Computer Science. 1994; 124-134, doi.org/10.1109/sfcs.1994.365700.
[10] Deutsch D, Jozsa R. Rapid solution of problems by quantum computation. Proceedings of the Royal Society of London. 439(1907). 1992;553-558, doi.org/10.1098/rspa.1992.0167.
[11] Petrenko A, Petrenko S, Gnidko K, Popova K, Grigorieva N. Application of quantum fourier transform in cryptology. Journal of Computer Virology and Hacking Techniques. 2025;21:26, doi.org/10.1007/s11416-025-00563-3
[12] Bernstein DJ. Introduction to post-quantum cryptography. Springer. 2009; https://link.springer.com/chapter/10.1007/978-3-540-88702-7_1.
[13] Nielsen MA, & Chuang IL. Quantum Computation and Quantum Information. Cambridge University Press. 2000; https://doi.org/10.1017/CBO9780511976667.
[14] Aspuru-Guzik A, Dutoi AD, Love PJ, Head-Gordon M. (2025). Simulated Quantum Computation of Molecular Energies. Science, 309(5741), 1704-1707. https://doi.org/10.1126/science.111347.
[15] Harrow A, Hassidim A, Lloyd S. Quantum algorithm for solving linear systems of equations. Physical Review Letters. 2009;103(15):150502, doi.org/10.1103/PhysRevLett.103.150502.
[16] Peruzzo A, McClean J, Shadbolt P, Yung MH, Zhou XQ, Love PJ, Aspuru-Guzik A, O'Brien JL. A variational eigenvalue solver on a photonic quantum processor. Nature Communications. 2014;5:4213, doi.org/10.1038/ncomms5213.
[17] McClean JR, Romero J, Babbush R, Aspuru-Guzik A. The theory of variational hybrid quantum-classical algorithms. New Journal of Physics. 2016;18:023023, doi.org/10.1088/1367-2630/18/2/023023.
[18] Cao Y, Romero J, Olson JP, Degroote M, Johnson PD, KieferováIan M, Kivlichan ID, Menke T, Peropadre B, Sawaya NPD, Sim S, Veis L, Aspuru-Guzik A. Quantum chemistry in the age of quantum computing. Chemical Reviews. 2019;119(19), doi.org/10.48550/arXiv.1812.09976.
[19] Rebentrost P, Mohseni M, Lloyd S. Quantum support vector machine for big data classification. Physical Review Letters. 2014;113:130503, doi.org/10.1103/PhysRevLett.113.130503.
[20] Biamonte J, Wittek P, Pancotti N, Rebentrost P, Wiebe N, Lloyd S. Quantum machine learning. Nature. 2017;549:195–202, doi.org/10.1038/nature23474.
[21] Havlíček V, Córcoles AD, Temme K, Harrow AW, Kandala A, Chow JM, Gambetta JM. Supervised learning with quantum-enhanced feature spaces. Nature. 2019;567:209–212,doi.org/10.1038/s41586-019-0980-2.
[22] Brassard G, Hoyer P, Mosca M, Tapp A. Quantum amplitude amplification and estimation. Quantum Computation and Quantum Information, Samuel J. Lomonaco, Jr. (editor), AMS Contemporary Mathematics. 2002;305:53-74, doi.org/10.1090/conm/305/05215
[23] Woerner S, Egger DJ. Quantum risk analysis. npj Quantum Information. 2019; 5(15), https://doi.org/10.1038/s41534-019-0130-6.
[24] Somma R, Somma RD, Boixo S, Barnum H, Knill E. Quantum simulations of Monte Carlo methods. Physics. Review Letters. 2008;101:130504, doi.org/10.1103/PhysRevLett.101.130504.
[25] Boneh D, Lipton R. Quantum crypto-analysis of hidden linear functions. Annual International Cryptology Conference. 1995;27, doi.org/10.1007/3-540-44750-4_34.
[26] Shor PW. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. SIAM Journal of Computing. 1997;26(5), doi.org/10.1137/S0097539795293172.
[27] Lau ASM, Tsui E, Lee WB. An Ontology-based Similarity Measurement for Problem-based Case Reasoning. Expert systems and Application. 2009;36(3): 6574-6579

[28] Lishman J. Qiskit: An open-source quantum computing framework. IBM GitHub repository. 2013, https://github.com/Qiskit/qiskit.
[29] Roffe J. Quantum Error Correction: An Introductory Guide. Contemporary Physics. 2019;60(3):226-245, doi.org/10.1080/00107514.2019.1667078.
[30] Jayashankar A, Mandayam P. Quantum Error Correction: Noise-adapted Techniques and Applications. Journal of Indian Institute of Science. 2022; doi.org/10.1007/s41745-022-00332-x.
[31] Cicero A, Maleki MA, Azhar MW, Kockum AF, Trancoso P. Simulation of Quantum Computers: Review and Acceleration Opportunities. ACM Transactions on Quantum Computing. 2025;7(1):1-35. https://doi.org/10.1145/3762672.
[32] Radojičić D, Kredatus S. The impact of stock market price Fourier transform analysis on the Gated Recurrent Unit classifier model. Expert Systems with Applications. 2020;159:113565, doi.org/10.1016/j.eswa.2020.113565.
[33] Stadnik B, Raudeliuniene J. Fourier Analysis for Stock Price Forecasting: Assumption and Evidence. Journal of Business Economics and Management. 2016; 17(3):365-380, doi.org/10.3846/16111699.2016.1184180.
[34] Dar GF, Padi TR, Sarode R, Dar QF. Stochastic Modeling for the Analysis and Forecasting of Stock Market Trend using Hidden Markov Model. Asian Journal of Probability and Statistics. 2022 ;18(1):43-56, doi.org/10.9734/AJPAS/2022/v18i130436.
[35] Elbatal I, Almetwally EM, Alaskar MZ, Alsaedi, BSO, Alzubaidi S, Yusuf M. Predictive analytics in accounting: Using regression models to forecast financial outcomes. Alexandria Engineering Journal. 2025;128:484-492, doi.org/10.1016/j.aej.2025.05.087.
[36] Kundu T, Pinsky E. Predicting daily stock price directions with deep learning models. Machine Learning with Applications. 2025;22:100744, doi.org/10.1016/j.mlwa.2025.100744.
[37] Osman EGA, Otaibi FA. Integrating deep learning and econometrics for stock price prediction: A comprehensive comparison of LSTM, transformers, and traditional time series models. Machine Learning with Applications. 2025;22: 100730, doi.org/10.1016/j.mlwa.2025.100730.
[38] Xu X, Zhang Y. Individual time series and composite forecasting of the Chinese stock index. Machine Learning with Applications. 2021,5:100035, doi.org/10.1016/j.mlwa.2021.100035.
[39] Basher SA, Sadorsky P. Forecasting Bitcoin price direction with random forests: How important are interest rates, inflation, and market volatility? Machine Learning with Application. 2022;9:100355, doi.org/10.1016/j.mlwa.2022.100355.
[40] Mostafavi SM, Hooman AR. Key technical indicators for stock market prediction. Machine Learning with Applications. 2025;20:100631, doi.org/10.1016/j.mlwa.2025.100631.
[41] McHugh C, Coleman S, Kerr D. Technical indicators for energy market trading. Machine Learning with Applications. 2021;6: 100182, doi.org/10.1016/j.mlwa.2021.100182.
[42] Fatouros G, Soldatos J, Kouroumali K, Makridis G, Kyriazis D. Transforming sentiment analysis in the financial domain with ChatGPT. Machine Learning with Applications. 2023 ;14:100508, doi.org/10.1016/j.mlwa.2023.100508.
[43] Lau ASM, Yen J, Chau PYK. Adoption of On-line Trading in Hong Kong Financial Market. Journal of Electronic Commerce Research. 2001;2(2):58-65.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261110

A Simulation-Based AI-Enabled Supply Chain Decision Framework for Cross-Border Business Expansion

Jiawei Qiu, Liyang Zeng[1], Zhuolun Wu, Chuyi Zou and Ben Fan
Guangdong University of Foreign Studies, Guangzhou, China

Abstract. This study develops an AI-enabled decision framework for cross-border supply chain management of small and medium-sized enterprises (SMEs). A SaaS-based platform integrates heterogeneous enterprise data into a unified supply chain state. We employ three task-oriented agents for demand forecasting, route planning, and data-value evaluation, and use an LLM-RAG component to provide semantic decision support for interpreting quantitative outputs. We conduct a 12-period multi-market simulation with four supply nodes and six demand markets, and compare the proposed framework with a traditional experience-driven baseline under identical conditions. The results indicate lower average MAPE and reduced inventory fluctuation under the proposed framework compared with the traditional baseline. The results provide preliminary evidence that the framework may improve demand forecasting accuracy and inventory stability under the simulated conditions.

Keywords. Artificial Intelligence (AI); Cross-border E-commerce; Intelligent Supply Chain; Decision Support Framework; Simulation Evaluation; Data Assetization

1. Introduction

AI-enabled supply chain systems have emerged as an important approach for improving decision support in cross-border business operations by integrating data analytics with supply chain management. The proposed approach aims to address decision-making challenges faced by small and medium-sized enterprises (SMEs) in demand forecasting, inventory management, and logistics planning, particularly under conditions of limited information availability. Based on a SaaS-based platform with LLM-RAG as a semantic decision-support component, this study does not establish its independent quantitative effect and makes three main contributions. First, it develops a

[1] Corresponding Author: Liyang Zeng, School of Economics and Trade, Guangdong University of Foreign Studies, China; Email: 2542747792@qq.com.

This study is financially supported by the Undergraduate Innovation Training Project of Guangdong University of Foreign Studies in 2026.

system-level framework integrating heterogeneous enterprise data into a unified dynamic supply-chain state. Second, it establishes a sequential decision mechanism linking demand forecasting with adaptive supply-chain configuration. Third, it evaluates the framework through multi-period simulations against conventional quantitative baselines under controlled conditions.

2. Research Status

Existing studies have examined AI applications in logistics optimization, supply chain resilience, demand forecasting, and intelligent decision support. Min [1] reviewed the theoretical and practical applications of AI in supply chain management. Chouhan et al. [2] highlighted the role of AI in logistics management, while Alikhani et al. [3] emphasized real-time perception, risk identification, and dynamic adjustment for supply chain viability. Jafari et al. [4] and Sharma et al. [5] further discussed generative AI and predictive analytics in supply chain management. Digital transformation and SaaS platforms have also been studied as mechanisms for data integration and coordination [6,7]. Digital supply chain twins have further been explored for dynamic supply chain management and resilience [8].

However, limited attention has been given to integrating data processing, forecasting, logistics planning, and decision support within a unified supply-chain framework. This study addresses this gap by linking data integration, demand forecasting, adaptive routing, and data-value evaluation within a unified dynamic state. Three questions are considered: how heterogeneous enterprise data can be integrated into a dynamic supply-chain state; how demand forecasts can be translated into adaptive supply-chain configuration; and how data-related value can be incorporated into decision processes. The main novelty is the coupling mechanism that links market perception to supply-chain reconfiguration and operational decisions to data-value evaluation.

3. Framework Design of an AI-Based Supply Chain System for Cross-Border Business Expansion

The proposed framework integrates multi-source enterprise data through a SaaS-based platform and employs three sequential agents: a demand forecasting agent, a supply-chain route-planning agent, and a data-value agent. The forecasting agent estimates one-period-ahead demand from historical sales, market trends, and consumer behavior using XGBoost. The route-planning agent combines predicted demand with inventory, supplier capacity, cost, logistics time, and risk to determine feasible allocations. The data-value agent evaluates the operational and potential value associated with data-driven decisions. The LLM-RAG component provides semantic decision support for interpreting quantitative outputs and does not independently determine quantitative decisions. The three agents therefore operate as a sequential decision process rather than as independent analytical functions. An LLM-RAG component based on Qwen2.5-7B-Instruct provides semantic decision support for interpreting quantitative outputs, while its independent quantitative effect is not

established in this study. Their outputs are connected through a unified supply-chain state, allowing changes in predicted demand to propagate to subsequent decisions.

To represent the enterprise as a dynamic system, the operational state at time t is defined as:

$$X_t=(M_t,P_t,S_t,I_t,L_t,D_t,C_t,R_t,V_t) \tag{1}$$

where the state includes market demand, sales, suppliers, inventory, logistics, consumer behavior, cost, risk, and data asset value. The market-demand and consumer-behavior components describe regional demand and relevant behavioral signals.

To evaluate data integration, a data collaboration index is defined as:

$$K_t = \frac{N_c}{N_a} \tag{2}$$

where the numerator is the number of effectively interconnected data modules and the denominator is the total number of modules. A higher index indicates stronger integration and provides a common data foundation for AI-enabled decisions.

The forecasting agent predicts future demand from historical sales, market trends, and consumer behavior using XGBoost:

$$\widehat{D_t}=f_{AI}(H_t,M_t,B_t) \tag{3}$$

where the inputs denote historical sales, market trends, and consumer behavior, and the forecasting function is parameterized by the model parameters. The supply-node and demand-node sets are defined as:

The comprehensive optimization objective identifies feasible supply-demand allocations that balance cost, logistics time, risk, and data value:

$$(S_i, E_j)^* = argmin(F) \tag{4}$$

Supply-chain cost, logistics time, and supply-chain risk are respectively defined as:

$$C=C_p + C_l + C_i \tag{5}$$

$$T=T_p + T_l \tag{6}$$

$$R=f(policy,market,transport) \tag{7}$$

Data value is defined as the combination of operational optimization value, model service value, and ecosystem value:

$$V=V_1 + V_2 + V_3 \tag{8}$$

In the simulation, operational optimization value is assessed only from explicitly measured cost and inventory components. Model service and ecosystem value are

treated as conceptual extensions unless corresponding revenues or economic benefits are explicitly simulated.

Figure 1 summarizes the evolutionary mechanism of the framework, linking data integration to demand perception, adaptive configuration, and data-value evaluation.

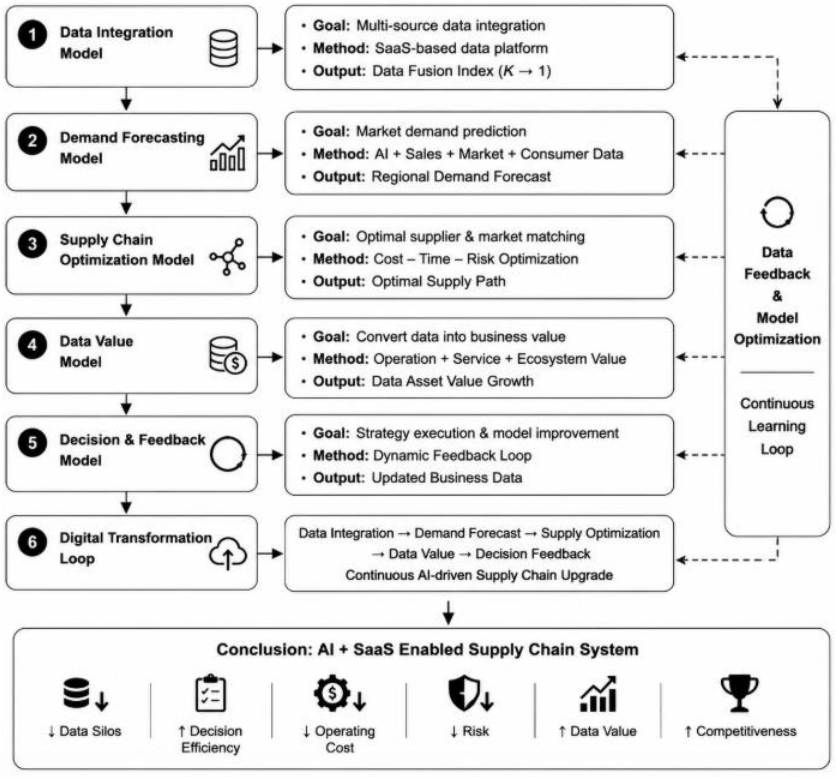


Figure 1. Evolutionary mechanism of the AI-based supply chain system framework for cross-border business expansion.

4. Experimental Results

This study evaluates the proposed AI-driven supply chain framework through multi-period simulation experiments designed to represent cross-border SME decision-making scenarios. Because the framework has not been validated through large-scale real-world deployment, simulation-based evaluation is used to examine its performance under controlled conditions. The 12-period horizon is selected to represent a multi-period decision process while maintaining a tractable simulation setting. The comparison uses the same simulation environment and 12-period decision horizon for the traditional experience-driven baseline and the proposed framework. Both approaches are evaluated under identical simulation conditions to ensure comparability. The evaluation focuses on demand forecasting accuracy, inventory stability, and order fulfillment performance.

Demand forecasting accuracy is evaluated using Mean Absolute Percentage Error (MAPE):

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{D_t - \widehat{D}_t}{D_t}\right| \times 100\% \quad (9)$$

MAPE measures the average absolute percentage deviation between predicted and actual demand over the evaluation periods. Where ndenotes the total number of decision-making periods, D_t represents the actual demand in period t, and $\widehat{D}_t$ represents the predicted demand. While MAPE is sensitive to very small actual-demand values, it is interpreted within the defined simulation demand conditions. A lower MAPE indicates a smaller forecasting deviation. In this study, MAPE is calculated consistently across the 12 decision-making periods.

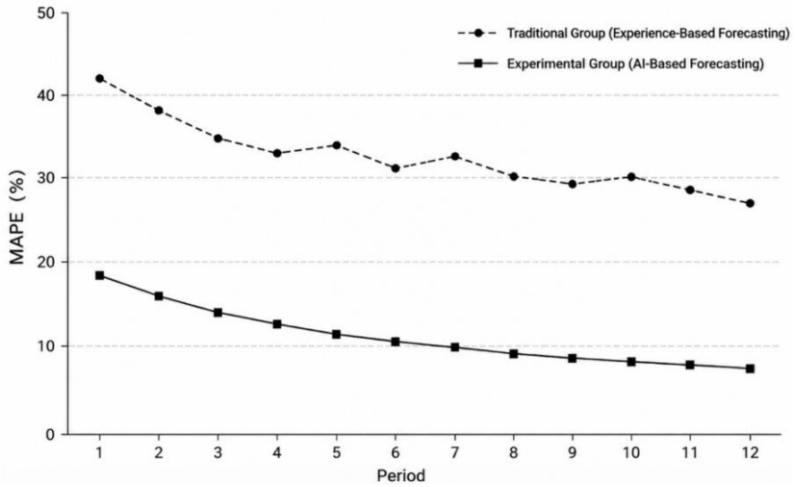


Figure 2. Comparison of MAPE for Demand Forecasting

Figure 2 compares the forecasting performance of the traditional experience-driven approach and the proposed AI-driven approach across the 12 decision-making periods. The proposed forecasting agent maintains lower MAPE values throughout the evaluation period, while both approaches show decreasing prediction errors over time. Across the 12 decision-making periods, the average MAPE decreases from 27.0% for the traditional baseline to 7.0% for the proposed framework. The lower forecasting error may provide more reliable demand information for subsequent inventory planning, procurement, and resource allocation under the simulated conditions.

Figure 3 compares order fulfillment performance. The proposed framework shows a higher and relatively stable order fulfillment rate, while the inventory fluctuation is lower than that of the traditional baseline. These results suggest potentially improved coordination between inventory stability and service performance under the simulated conditions.

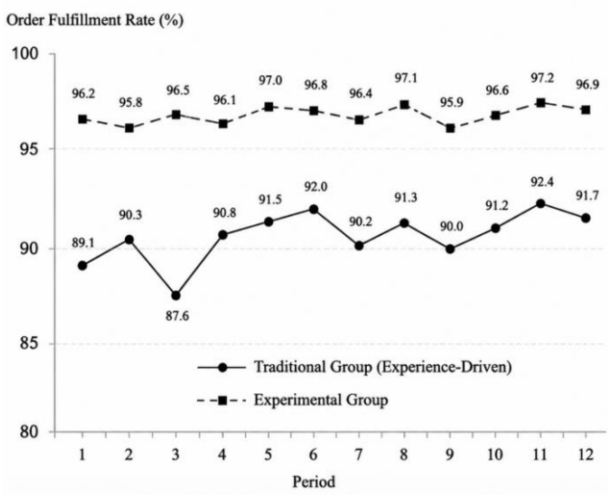


Figure 3. Comparison of Order Fulfillment Rate.

Overall, the simulation results provide preliminary evidence that the integrated framework can improve demand forecasting accuracy, inventory stability, and fulfillment performance under the defined experimental conditions. These findings should be interpreted as simulation-based evidence rather than direct evidence of generalizable real-world performance. Taken together, the results presented in Figures 2 and 3 indicate that the benefits of the proposed AI-driven framework are not limited to forecasting accuracy. More accurate demand information can be transmitted to subsequent inventory and fulfillment decisions, allowing the system to adjust inventory levels while maintaining service performance. The reduction in inventory fluctuation, together with the improvement in order fulfillment, therefore provides simulation-based evidence of more coordinated decision-making across the supply chain.

5. Conclusion

Traditional cross-border supply chain management often relies on experience-driven decisions and fixed operational processes, which may lead to forecasting errors, inventory imbalance, and delayed responses under changing market conditions.

The AI-enabled supply-chain framework proposed in this study integrates heterogeneous data, demand forecasting, adaptive route planning, and data-value evaluation, with LLM-RAG as a semantic decision-support layer, to support dynamic supply-chain decisions. Simulation results indicate lower average MAPE and reduced inventory fluctuation under the proposed framework compared with the traditional experience-driven baseline. These results provide preliminary evidence of potentially improved forecasting accuracy, inventory stability, and fulfillment performance under the defined simulation conditions. However, further validation using real-world operational data is required.

References

[1] Min H. Artificial intelligence in supply chain management: Theory and applications. International Journal of Logistics Research and Applications. 2024;27(2):155-178. doi:10.1080/13675567.2023.2181234.
[2] Chouhan V, Saraswat P, Goswami S. AI's role in efficient logistics management: shaping the future of supply chain in India. Transportation Research Interdisciplinary Perspectives. 2026;39:102149-102149. doi:10.1016/J.TRIP.2026.102149.
[3] Alikhani R, Hosseinzadeh M, Agrawal N. Enhancing supply chain viability with artificial intelligence:A forward-thinking perspective.Annals of Operations Research. 2026;(prepublish):1-49.doi:10.1007/S10479-026-07349-W.
[4] Jafari M, Akbari HA, Akhavan P. Leveraging generative AI to drive supply chain resilience and sustainable performance: A mediation framework for next-generation supply chains. Technology in Society. 2026;88:103444-103444. doi:10.1016/J.TECHSOC.2026.103444.
[5] Sharma P, Gunasekaran A, Subramanian G. Enhancing Supply Chain: Exploring and Exploiting AI Capabilities. The Journal of Computer Information Systems. 2026;66(4):391-405. doi:10.1080/08874417.2024.2386527.
[6] Kamble SS, Gunasekaran A, Sharma R. Modeling the blockchain enabled traceability in agriculture supply chain. International Journal of Information Management. 2020;52:101967. doi:10.1016/j.ijinfomgt.2019.05.023.
[7] Bag S, Gupta S, Kumar A, Sivarajah U. Role of artificial intelligence in supply chain management and organizational performance. International Journal of Production Research. 2021;59(16):4902-4951. doi:10.1080/00207543.2020.1788738.
[8] Ivanov D, Dolgui A. A digital supply chain twin for managing the disruption risks and resilience in the era of Industry 4.0. Production Planning & Control. 2021;32(9):775-788. doi:10.1080/09537287.2020.1768450.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)

© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261111

Research on Evaluation of Large Language Models for Healthcare

Peng Cheng[a], Hong Ruan[a], Xin Liao[a] and Mingliang Tan [b,1]
[a] *Nanchong Gaoping District People's Hospital, Nanchong 637100, China*
[b]*Key Laboratory of Digital-Intelligent Disease Surveillance and Health Governance, North Sichuan Medical College, Nanchong 637000, China*

Abstract. Large language models (LLMs) constitute a new growth engine for artificial intelligence, boasting tremendous potential to reshape the healthcare industry. When deploying LLMs to empower healthcare scenarios, scientific and comprehensive evaluation is an indispensable task, which can effectively mitigate potential risks and guarantee service quality. This paper elaborates on the evaluation contents and methodologies for LLMs oriented toward healthcare, and identifies key research priorities for future exploration. The work provides references for academic research and industrial practice concerning the evaluation of medical LLMs, enabling LLMs to exert greater positive impacts across the healthcare sector.

Keywords. healthcare; large language model evaluation; ChatGPT; generative artificial intelligence

1.Introduction

Empowering the healthcare industry with new-generation information technologies including artificial intelligence, big data, cloud computing, mobile internet and the Internet of Things carries profound significance. As one of the most groundbreaking and representative research achievements in artificial intelligence in recent years, large language models are deep learning architectures trained to comprehend and generate natural language, exemplified by OpenAI's ChatGPT, Google's PaLM 2, Alibaba's Tongyi Qianwen, Baidu's ERNIE Bot, etc.

Healthcare stands as one of the sectors most profoundly influenced by LLMs. LLMs can be integrated into numerous links and scenarios of medical services, serving diverse stakeholders including clinicians, nurses, patients, medical educators, the general public, disease control specialists and scientific researchers. Such integration accelerates intelligent transformation and high-quality development of the healthcare industry, advancing the Healthy China Initiative. To minimize risks, ensure service quality and maximize the positive value of LLMs, systematic, scientific and comprehensive evaluation of medical LLMs emerges as a critical and mandatory undertaking, attracting widespread attention from both academia and industry.

[1] Corresponding Author: Mingliang Tan, Key Laboratory of Digital-Intelligent Disease Surveillance and Health Governance, North Sichuan Medical College, Nanchong 637100, China; Email: tanml@whu.edu.cn.

2.Related works

Existing research primarily evaluates medical LLMs by leveraging medical licensing examinations, specialized medical question banks and medical QA datasets. For instance, Kung et al. assessed ChatGPT's performance on the United States Medical Licensing Examination (USMLE) and found that ChatGPT could reach or nearly reach the passing threshold without domain-specific training or reinforcement learning [1]. Alfertshofer et al. further tested ChatGPT against medical licensing exams from six jurisdictions: the United States, Italy, France, Spain, the United Kingdom and India, revealing substantial cross-nation accuracy discrepancies. ChatGPT achieved its highest accuracy of 73% on Italian exam questions, while scoring merely 22% (the lowest) on French examination items [2]. Beyond multiple-choice licensing tests, Nori et al. explored GPT-4's performance on complex medical challenge problems and pointed out that standard exam benchmarks cannot fully reflect real-world clinical reasoning capacity [3]. Singhal et al. also demonstrated that even high-performing LLMs may exhibit unstable performance when facing out-of-distribution medical questions derived from real clinical workflows [4].

Scholars also design evaluation prompts based on disease prevention, diagnosis, treatment and nursing guidelines, clinical experience and real patient cases to assess medical LLMs. Sarraju et al. constructed a test set of 25 questions covering risk factors, laboratory results and medication information, and evaluated ChatGPT against this dataset; the model delivered appropriate responses to 21 out of 25 questions (84%) [5]. Walker et al. restated clinical guideline recommendations from five major disease categories into test prompts and input them to ChatGPT. Adopting a modified Evaluation of Quality of Information for Patients (EQIP) instrument, their research indicated a 60% consistency rate between ChatGPT outputs and official guideline advice, and that the medical information quality produced by ChatGPT was comparable to static online medical resources [6]. Tang et al. conducted systematic evaluation of LLMs on medical-evidence summarization tasks and pointed out that LLMs still suffer factual inconsistency when summarizing biomedical evidence, which requires careful human expert verification for open-ended medical outputs [7]. Xie et al. built DoctorFLAN, a large-scale real-workflow-oriented medical dataset, and emphasized that evaluation prompts should align with actual clinical work tasks rather than purely academic examination questions [8].

Multiple studies have benchmarked LLMs developed by different vendors and with different versions for medical tasks. Koga et al. tested ChatGPT-3.5, ChatGPT-4 and Google Bard using clinical data from 25 patients with neurodegenerative diseases, and confirmed that ChatGPT-4 delivered the highest diagnostic accuracy among the three models [9]. Mallio et al. compared four LLMs (ChatGPT-3.5, GPT-4, Perplexity and Bing) in generating structured radiology reports, analyzing their performance and application potential [10]. Sandmann et al. performed head-to-head benchmarking among multiple open-source and closed-source medical LLMs under identical clinical decision-making scenarios and observed significant performance gaps between general-purpose models and domain-optimized variants [11]. Ouyang et al. presented CliMedBench, a large-scale Chinese clinical benchmark covering 14 real-world clinical scenarios, which provides standardized pipelines for fair and reproducible cross-model comparison of medical LLMs [12].

Apart from answer accuracy measured by examinations and clinical case prompts, recent research has highlighted safety, bias and hallucination risks when deploying

LLMs in healthcare scenarios. Omiye et al. demonstrated that large language models can inherit racial bias embedded within training corpora, which poses potential risks for clinical decision-support applications [13]. To systematically measure such safety defects, domain-specific medical evaluation benchmarks have been developed. Bedi et al. proposed MedHELM, a multi-task benchmark suite supporting holistic evaluation of LLMs across diverse clinical tasks [14]. Targeting the frequent medical hallucination problem, Pandit et al. built MedHallu, a dedicated benchmark for detecting factual errors in medical LLM outputs [15]. Nevertheless, existing evaluation practices still exhibit obvious limitations. Du et al. conducted a systematic review on generative LLM evaluation for EHR-related clinical applications and emphasized that most existing assessments concentrate merely on final-stage output performance, while overlooking upstream critical links covering data source, model training and real-world deployment [16]. Han et al. performed a systematic review covering LLM evaluation studies in clinical medicine and pointed out that evaluation outcomes obtained from public static benchmarks often cannot reliably predict model performance in real-world clinical environments [17].

Overall, evaluating LLMs for medical and healthcare applications has become a hot interdisciplinary research topic spanning medicine, nursing, computer science and information management, with rapidly growing publications in top domestic and international academic journals. Current literature predominantly focuses on testing the practical performance and application potential of LLMs represented by ChatGPT within medical scenarios, yet lacks macro-level systematic discussions on the full spectrum of evaluation contents and methodologies, which serve as the prerequisite and foundation for credible LLM assessment. This paper addresses this research gap, offering actionable references for academic exploration and industrial deployment of medical LLM evaluation, so as to advance LLM-enabled medical services.

3.Methods

This study adopts a literature review combined with a full-lifecycle decomposition research framework to construct a complete evaluation system for large language models oriented to medical and healthcare scenarios. Five academic databases including PubMed, Scopus, Web of Science, IEEE Xplore, and arXiv were retrieved, covering clinical medical publications, artificial-intelligence conference papers, and open-access preprint manuscripts to comprehensively collect research related to the evaluation of medical large language models. The Boolean search query `(("large language model" OR "LLM" OR "ChatGPT" OR "GPT-4") AND ("medical" OR "healthcare" OR "clinical") AND ("evaluation" OR "benchmark" OR "assessment" OR "testing"))` was applied to title, abstract and keyword fields, with the literature retrieval cutoff date set to March 2026, covering the main development period of medical-oriented large-language-model evaluation research.

The inclusion criteria are as follows: original empirical research, conference papers, or preprint manuscripts focusing on performance testing, benchmark construction, or evaluation framework exploration for large language models in medical and clinical scenarios; studies that report specific test datasets, evaluation indicators, experimental results, or practical assessment pipelines for medical LLMs; full-text accessible resources written in English. Exclusion criteria include pure descriptive reviews, editorials, letters, commentaries without empirical evaluation experiments; research

focusing on traditional non-LLM NLP models or pure multimodal models without pure-text medical LLM evaluation content; articles only discussing application prospects of medical artificial intelligence, without any evaluation-related experiments or analysis; duplicate publications, documents with incomplete metadata and inaccessible full-text content. All retrieved literature records were imported into reference-management software for duplicate removal. Two researchers independently performed two-round screening: title-abstract screening first, followed by full-text assessment for potentially eligible articles. Any disagreements about inclusion status were resolved through group discussion among researchers. After the final screening, we extracted key information from included papers, including evaluation tasks, benchmark datasets, evaluation metrics, experimental conclusions, and existing research limitations.

On the basis of literature analysis and domain knowledge, the complete industrial chain of medical LLM research and application is divided into eight sequential links: data collection, data processing, model selection, model training, model testing, model tuning, model deployment, and model application. A multi-dimensional evaluation module system covering the whole lifecycle of medical LLMs is constructed. Furthermore, this paper sorts out seven mainstream operable evaluation methodologies suitable for medical scenarios, as shown in Figure 1. This mixed-method framework can balance evaluation efficiency, objectivity and domain professionalism, and support comprehensive and credible assessment of medical LLMs under complex real-world healthcare conditions.

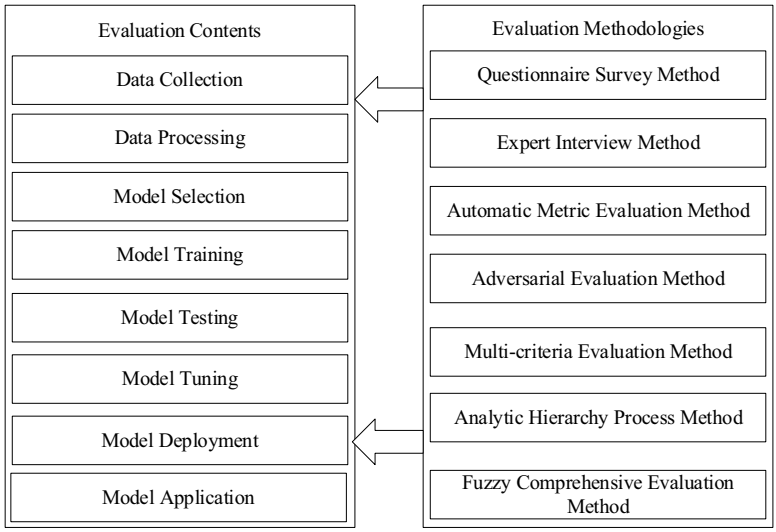


Figure 1. A Multi-dimensional Evaluation Module System of Large Language Models for Healthcare.

4.Results & Discussion

4.1 Evaluation Contents of Large Language Models for Healthcare

From the full lifecycle of LLM construction and deployment, this paper divides the evaluation framework of medical LLMs into eight interconnected modules: data collection evaluation, data processing evaluation, model selection evaluation, model

training evaluation, model testing evaluation, model tuning evaluation, model deployment evaluation and model application evaluation, enabling holistic, systematic and in-depth assessment.

(1) Data collection evaluation. Training medical LLMs relies on massive multi-source datasets derived from professional monographs, academic journals, online health communities, electronic medical records, health questionnaires and news media. Data collection evaluation covers whether acquired data aligns with the demands of medical domains and specific business workflows, alongside comprehensive assessment of data source credibility, volume, integrity, security, consistency, representativeness, diversity, time coverage and annotation costs. Additionally, evaluators must verify whether data collection complies with privacy and cybersecurity regulations, avoids intellectual property infringement, and adheres to medical ethics codes and standards.

(2) Data processing evaluation. Data processing constitutes a complex, pivotal stage of LLM training; the adopted workflows, technologies and schemes directly determine training feasibility and downstream model performance. Data processing evaluation addresses the following dimensions: proportional balance across heterogeneous data categories; rationality of techniques for identifying and resolving duplicate, missing, abnormal and inconsistent entries; necessity of data synthesis and augmentation to expand dataset diversity and boost model generalization; compliance of data formats and structures with model input specifications; efficiency of data preprocessing pipelines; and appropriateness of data storage infrastructure.

(3) Model selection evaluation. Deep neural network architectures form the core foundation for building medical LLMs. Mainstream medical LLMs are built on the Transformer architecture, while Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and Gated Recurrent Units (GRUs) serve as alternative architectures for traditional NLP. Model selection evaluation analyzes candidate architectures from multiple angles: whether their performance, computational efficiency and task strengths match medical information service requirements; hardware resource thresholds; flexibility and scalability; availability of pre-trained weights for transfer learning and fine-tuning; and supported programming languages.

(4) Model training evaluation. Prior to formal training, models undergo parameter initialization via random initialization or pre-trained weight transfer to facilitate adaptation to training corpora. Large-scale text corpora are then fed into neural networks in batches, with backpropagation algorithms adjusting model parameters to minimize loss functions. Model training evaluation assesses: whether hyperparameter configurations (learning rate, batch size, epoch count, regularization strength) and optimizer selection positively impact model performance; sufficiency of hardware and software infrastructure to sustain deep network training; acceptability of total training duration; and efficiency of data and computational resource utilization.

(5) Model testing evaluation. Two core pillars underpin medical LLM testing: construction of domain-specific benchmark datasets, and selection of task-aligned evaluation metrics. Model testing evaluation focuses on: time and economic costs of dataset construction; reusability of public medical benchmarks such as MedQA-USMLE, MedMCQA and PubMedQA; dataset quality, scale, coverage and train/validation/test partitioning schemes; suitability of quantitative metrics including accuracy, recall, F1-score, perplexity, BLEU, ROUGE and METEOR; and comprehensive testing of model safety, reliability and fairness.

(6) Model tuning evaluation. LLM tuning spans data preprocessing, architectural design, hyperparameter adjustment, regularization, optimizer/loss function design,

ensemble learning, knowledge distillation and hardware configuration, requiring iterative trials to elevate model performance, efficiency and adaptability. Model tuning evaluation comprehensively assesses all above dimensions; for network architecture optimization, key inspection items include layer count, neuron quantity and activation function selection. Notably, model testing and tuning form an iterative closed loop: post-testing tuning precedes subsequent re-testing cycles.

(7) Model deployment evaluation. LLM deployment relies on five critical dimensions: hardware resources, software environments, network bandwidth, data security and scalability. Massive computational resources are required for LLM training and inference, mandating well-matched hardware configurations for stable, efficient runtime. Supporting software infrastructure encompasses operating systems, deep learning frameworks and parallel computing engines. Sufficient network bandwidth eliminates transmission bottlenecks and accelerates service response. Robust safeguards must be implemented to preserve data privacy and security, while scalable architecture facilitates future model expansion and upgrades. Deployment evaluation rigorously audits the rationality of all five dimensions.

(8) Model application evaluation. Technical development and model construction are not end goals; LLMs are ultimately deployed to real-world scenarios to resolve practical challenges, streamline workflows, innovate service paradigms, improve public well-being and drive social progress. Model application evaluation analyzes integrated LLM systems across the following criteria: accuracy, authority, coherence and readability of generated medical outputs; intuitiveness, aesthetics and ergonomics of user interfaces; system compatibility, responsiveness and stability (avoiding excessive latency, stuttering or crashes); provision of traceable reference citations for medical claims; and presence of feedback collection pipelines to iteratively refine output quality.

4.2 Evaluation Methodologies of Large Language Models for Healthcare

This section addresses the operational approaches for LLM assessment, covering seven mainstream methodologies: questionnaire survey, expert interview, automatic metric evaluation, adversarial evaluation, multi-criteria evaluation, Analytic Hierarchy Process (AHP) and fuzzy comprehensive evaluation. Each method carries distinct advantages and limitations; integrated deployment of multiple methodologies enables holistic, in-depth model assessment and delivers actionable insights for model optimization.

(1) Questionnaire survey method. The questionnaire survey method collects user feedback via standardized pre-designed questionnaires to quantify real-world LLM performance and identify practical defects. Key implementation principles include: developing comprehensive questionnaires incorporating open-ended items tailored to target scenarios and user groups; sampling representative respondents with sufficient sample size; and conducting reliability and validity tests to guarantee credible, consistent survey outcomes.

(2) Expert interview method. The expert interview method engages domain specialists in clinical medicine and artificial intelligence for in-depth discussions, leveraging professional expertise to capture nuanced assessments of LLM strengths, weaknesses and overall performance. Critical implementation considerations: recruiting experts with robust domain knowledge, industry experience and rigorous professional ethics; defining targeted evaluation dimensions prior to interviews to focus discussion scope; and adopting open dialogue formats rather than rigid closed questioning.

(3) Automatic metric evaluation Method. The automatic metric evaluation method objectively quantifies text generation quality via numerical metrics such as fluency, semantic relevance, factual consistency, accuracy, F1-score, BLEU, ROUGE and perplexity. It delivers rapid, reproducible cross-model comparison without subjective human annotation, boosting evaluation consistency. Limitations include susceptibility to training data bias, weak robustness against adversarial inputs, and limited capacity to measure output diversity.

(4) Adversarial evaluation method. The adversarial evaluation method constructs adversarial test samples to benchmark model robustness and safety against perturbed, malicious or disruptive inputs, identifying latent security vulnerabilities and ensuring stable performance under real-world interference. Common adversarial strategies include gradient-based attacks, transfer attacks, confidence attacks, decision attacks and ensemble attacks; practitioners select appropriate techniques aligned with task characteristics and application constraints.

(5) Multi-criteria evaluation method. The multi-criteria evaluation method synthesizes heterogeneous indicators to conduct rigorous, holistic LLM assessment. Core operational steps: deploying diversified metrics (accuracy, recall, F1-score, generalization capacity, etc.); rationally weighting indicators according to scenario-specific priorities; ensuring evaluation objectivity via controlled comparative experiments and cross-validation; and visualizing evaluation results through charts for intuitive cross-model comparison.

(6) Analytic hierarchy process method. AHP constitutes a qualitative-quantitative systematic hierarchical analysis technique for complex multi-objective, multi-criteria decision-making problems. Its core logic decomposes intricate evaluation challenges into layered sub-problems, quantifies relative importance via pairwise expert comparisons, and derives weighted evaluation standards. Standard implementation workflows: establishing hierarchical structural models reflecting logical element relationships; constructing pairwise comparison judgment matrices; single-level sorting and consistency testing; total hierarchical sorting and consistency testing; and deriving final evaluation conclusions.

(7) Fuzzy comprehensive evaluation method. Rooted in fuzzy mathematics, the fuzzy comprehensive evaluation method quantifies ambiguous, hard-to-measure indicators and delivers integrated assessments for vague evaluation criteria with unclear boundaries. Standard implementation steps: constructing factor sets of evaluation dimensions; establishing grading comment sets; calculating weight vectors; generating fuzzy relation matrices; conducting fuzzy synthesis operations; and interpreting comprehensive evaluation outputs. Its core strength lies in handling uncertainty and fuzziness in subjective evaluation tasks, yielding results that better reflect real-world conditions. Drawbacks include heavy reliance on expert judgment and elevated computational complexity.

5.Conclusion & Limitations

Deploying disruptive LLM technology within healthcare industries delivers transformative value: reshaping medical service paradigms, elevating patient experience and consultation efficiency, reducing patients' financial burdens, alleviating medical resource shortages, and mitigating uneven distribution of healthcare resources. To fully

unlock the positive value of LLMs in medical workflows while minimizing associated risks and adverse impacts, systematic, scientific, comprehensive and effective evaluation is mandatory.

This paper elaborates on the full spectrum of evaluation contents and methodologies for medical LLMs, providing macro-level theoretical and practical references for domain-specific LLM assessment. Future research priorities are outlined as follows: constructing standardized indicator systems for evaluating LLMs across diverse medical business scenarios; developing open-source libraries and dedicated evaluation platforms for automated medical LLM testing; exploring integrated frameworks that coordinate multiple evaluation methodologies to achieve globally optimized assessment performance.

This study has three main limitations. First, the proposed evaluation framework remains a theoretical conceptual construct and has not yet been empirically validated through real-world projects. Second, the literature search was restricted to English-language publications, excluding Chinese-language studies, which may introduce a language-based selection bias in the literature perspective. Third, this study focused exclusively on text-based medical large language models and did not encompass multimodal medical large language models.

Acknowledgments

This paper was supported by The Nanchong Social Science Research Project (No. NC26C393), Key Laboratory of Digital-Intelligent Disease Surveillance and Health Governance Project (No. JBSZ25-Y-09), Healthy Nanchong Research Center Project (No. NC26JK11), North Sichuan Medical College Special Research Cooperation Project (No. CBY25-ZXB14).

References

[1] Kung T H, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models[J]. PLoS digital health, 2023, 2(2): e0000198.

[2] Alfertshofer M, Hoch C C, Funk P F, et al. Sailing the seven seas: a multinational comparison of ChatGPT's performance on medical licensing examinations[J]. Annals of Biomedical Engineering, 2024, 52(6): 1542-1545.

[3] Nori H, King N, McKinney S M, et al. Capabilities of gpt-4 on medical challenge problems[J]. arXiv preprint arXiv:2303.13375, 2023.

[4] Singhal K, Azizi S, Tu T, et al. Large language models encode clinical knowledge[J]. Nature, 2023, 620: 172-180.

[5] Sarraju A, Bruemmer D, Van Iterson E, et al. Appropriateness of cardiovascular disease prevention recommendations obtained from a popular online chat-based artificial intelligence model[J]. JAMA, 2023, 329(10): 842-844.

[6] Walker H L, Ghani S, Kuemmerli C, et al. Reliability of Medical Information Provided by ChatGPT: Assessment Against Clinical Guidelines and Patient Information Quality Instrument[J]. Journal of Medical Internet Research, 2023, 25: e47479.

[7] Tang L, Sun Z, Idnay B, et al. Evaluating large language models on medical evidence summarization[J]. NPJ Digital Medicine, 2023, 6(1):158.

[8] Xie W, Xiao Q, Zheng Y, et al. Enabling doctor-centric medical AI with LLMs through workflow-aligned tasks and benchmarks[J]. NPJ Health Systems, 2025, 2(1):44.

[9] Koga S, Martin N B, Dickson D W. Evaluating the performance of large language models: ChatGPT and Google Bard in generating differential diagnoses in clinicopathological conferences of neurodegenerative disorders[J]. Brain Pathology, 2023: e13207.

[10] Mallio C A, Sertorio A C, Bernetti C, et al. Large language models for structured reporting in radiology: performance of GPT-4, ChatGPT-3.5, Perplexity and Bing[J]. La radiologia medica, 2023, 128: 808-812.
[11] Sandmann S, Hegselmann S, Fujarski M, et al. Benchmark evaluation of DeepSeek large language models in clinical decision-making[J]. Nature medicine, 2025, 31(8): 2546-2549.
[12] Ouyang Z, Qiu Y, Wang L, et al. CliMedBench: A large-scale Chinese benchmark for evaluating medical large language models in clinical scenarios[C]//Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing. 2024: 8428-8438.
[13] Omiye J A, Lester J C, Spichak S, et al. Large language models propagate race-based medicine[J]. NPJ Digital Medicine, 2023, 6: 195.
[14] Bedi S, Cui H, Fuentes M, et al. Holistic evaluation of large language models for medical tasks with MedHELM[J]. Nature Medicine, 2026, 32(5): 943-951.
[15] Pandit S, Xu J, Hong J, et al. Medhallu: A comprehensive benchmark for detecting medical hallucinations in large language models[C]//Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing. 2025: 2858-2873.
[16] Du X, Zhou Z, Wang Y, et al. Testing and evaluation of generative large language models in electronic health record applications: a systematic review[J]. Journal of the American Medical Informatics Association, 2026, 33(3): 743-753.
[17] Han S, Kim J, Lee J. A systematic review of large language model (LLM) evaluations in clinical medicine[J]. Journal of Medical Artificial Intelligence, 2025,7:18.

Section 3

Special Session on “New Ecology of Lifelong Education”

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)

© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261113

Mutual Information Analysis of Employment-Course Score Relationships for Curriculum Optimization in Higher Vocational Traditional Chinese Pharmacy

Ziquan Hu[a,1] and Shuhao Hu[b]

[a] *Informationization Construction and Service Center, Dazhou Vocational College of Chinese Medicine, Dazhou 635000, Sichuan, P. R. China*
[b]*School of Computer Science and Software Engineering, Southwest Petroleum University, Shunqing Dist, Nanchong 637001, Sichuan, P. R. China*

Abstract. Against rising graduate employment pressure and the widespread application of digital technology across industries, the curriculum of higher vocational Traditional Chinese Pharmacy (TCP) fails to fully match the competency requirements of actual jobs. This research intends to quantitatively analyze the correlation between students' course scores and employment performance, offering data support for curriculum revision and optimization in vocational colleges. Using data from 236 TCP graduates, the study classified employment destinations into six types and converted them into binary employment variables. Mutual Information (MI) was employed to quantify the nonlinear correlation between 49 course scores and employment results. Combined with a Monte Carlo-assisted elimination algorithm, the research filtered out courses with no statistical connection to employment. Repeated random sampling ensured stable numerical results and course rankings. The findings show that core TCP professional courses do not occupy leading positions in the MI ranking. Sensitivity tests by removing average grades further verified the reliability of the conclusions. This study proves MI is a valid tool for exploring links between courses and employment, while pointing out its limitation: MI only reflects correlation instead of causality. It is recommended that curriculum teams refer to MI rankings, align course design with talent training goals and industry needs, and add digital competency courses including artificial intelligence and big data analysis to adapt to industrial upgrading.

Keywords. Traditional Chinese Pharmacy; Higher Vocational Education; Mutual Information; Educational Data Mining; Curriculum Reform; Employment Outcomes

[1] Corresponding Author: Ziquan Hu, E-mail: ziquanhu@qq.com.

1. Introduction

Against the backdrop of escalating graduate employment pressure and the deep integration of digital technologies into the sustainable economy, higher vocational TCP education faces a persistent structural dilemma: the curriculum is organized by disciplinary logic, while industrial posts require competence logic. A substantial gap exists between existing curriculum content and actual industrial talent demands.

To inform curriculum review, a data-driven approach is needed to quantify associations between course scores and employment outcomes. Mutual information (MI) measures statistical dependence between variables without assuming linearity or specific distributions [1]. It can identify courses whose scores share the strongest statistical information with employment outcomes. However, MI only reflects statistical correlation rather than causal relationship, and it cannot automatically eliminate confounding interference such as students' general academic ability, inconsistent grading standards and teachers' teaching differences.

This study applies MI analysis to 49 course scores and employment outcomes of 236 TCP graduates. A Monte Carlo-aided elimination algorithm is adopted to filter out irrelevant courses with zero MI values, and multiple repeated trials are used to realize stable convergence of course correlation ranking results. Two supplementary analyses a sensitivity analysis removing the grade average feature, are included to assess robustness. The results show that TCP-specific courses are not overrepresented among the top-ranked courses. This finding is descriptive rather than prescriptive; it does not mean professional core courses are unimportant, but only indicates that their statistical correlation with binary employment indicators is weak under the current variable setting. Multiple alternative influencing factors such as employment signal effect, inter-course grading difference and general learning ability interference are further explained in the discussion section.

The study makes three contributions. First, it provides a replicable method for quantifying course-employment associations in vocational education using MI with Monte Carlo stabilization. Second, it reports empirical findings from a specific institutional context that may inform local curriculum review. Third, it explicitly delineates the boundaries of what MI can and cannot claim, offering a cautious template for similar data-driven educational analyses. This study does not claim pioneering innovation, and all research conclusions are objectively limited by sample scale and variable measurement methods.

2. Materials and methods

2.1 Study Design and Participants

The study was reviewed and approved by the Institutional Review Board of Dazhou Vocational College of Chinese Medicine. Academic and employment record linkage was conducted by designated institutional administrators from the Academic Affairs Office and Admissions and Employment Office under an official data sharing

agreement approved by the college administration. All raw data were fully de-identified beforehand; sensitive personal information including student names, ID numbers and class details was replaced with anonymous coding to strictly protect individual privacy and comply with standard academic research ethics. Only author was granted access to the de-identified linked dataset, and no personally identifiable information is presented in this paper. This study involved only routine educational record linkage without experimental interventions or research on vulnerable populations.

The dataset includes course scores and employment outcomes for 236 graduates (125 in 2022, 111 in 2023).

Table 1 Employment outcomes (five positive categories, one negative)

Category	N (%)
Signed employment agreement	98 (41.53%)
Signed labor contract	52 (22.03%)
Other official employment certification	11 (4.66%)
Grassroots employment (health commission, hospital, health center)	45 (19.07%)
Junior college upgrade to undergraduate programs	10 (4.24%)
Unemployed (job seeking)	20 (8.47%)
Overall employment and further education rate	91.53%

Primary outcome variable (binary): 0=unemployed (job seeking); 1=employed or further education (any of the five positive categories). This study adopts binary classification limited by sample size and data accessibility, which aggregates undergraduate admission, grassroots clinical employment and digital emerging posts into the same positive category, resulting in inability to distinguish the differential correlation between course performance and different employment paths. This inherent defect of variable setting is fully recognized and discussed in the subsequent limitations.

The curriculum includes 49 courses across seven modules: Ideological Quality (6 courses), Humanistic Science and Technology Quality (14 courses), Public Elective (2 courses), Professional Basic (12 courses), Professional Core (5 courses), Professional Expansion (7 courses), and Professional Skills Training (2 courses). The current curriculum lacks courses in AI, big data, or advanced digital literacy.

2.2 Mutual information

Mutual information measures the amount of information one random variable contains about another [1]. For two discrete random variables X (course score) and Y (employment status), their mutual information $I(X;Y)$ is defined as:

$$I(X;Y) = \sum_{x\epsilon X}\sum_{y\epsilon Y} p(x,y)\log\frac{p(x,y)}{p(x)p(y)},$$

where $p(x, y)$ is the joint probability distribution of X and Y, $p(x)$ and $p(y)$ are marginal probability distributions for X and Y, respectively. The value range of mutual information is $[0, \min\{H(X), H(Y)\}]$, where $H(X)$ and $H(Y)$ are the entropies of random variables X and Y, respectively. $I(X; Y) = 0$ if and only if X and Y are independent.

A key advantage of mutual information is that it is non-parametric and model-free, requiring no predefined assumption on data distribution. It is particularly suitable for educational score data characterized by unknown or non-normal

distributions, heterogeneous grading criteria, and mixed discrete-continuous attributes. College course scores have obvious non-normal distribution characteristics, and there are complex correlation relationships between different courses; MI can effectively adapt to such data characteristics and capture potential nonlinear associations.

In contrast, logistic regression and ordinal regression depend on rigid statistical assumptions that are rarely fully satisfied in course score datasets:

(1) Logistic regression assumes linearity in the logit, independent observations, absence of severe multicollinearity, and no complete separation;

(2) Ordinal regression additionally requires the proportional odds assumption, which is commonly violated given the complex hierarchical relationship between course performance and employment outcomes;

(3) Both regression approaches are sensitive to score distribution patterns, grading inconsistency, instructor-specific effects, and inter-course correlations, all of which exist in the present sample.

Mutual information effectively avoids these constraints and enables direct detection of nonlinear associations without distributional restrictions, making it more appropriate for exploratory analysis of course-employment relationships. In follow-up research, logistic regression and other traditional models will be introduced for comparative verification to further confirm the applicability advantages of MI method.

In this study, mutual information is estimated via sklearn.feature_selection.mutual_info_classif(). The method is capable of identifying courses with zero mutual information that are statistically independent of employment outcomes. Nevertheless, inherent randomness in the k-nearest neighbor entropy estimator and stochastic noise processing for tied scores lead to unstable mutual information values and inconsistent zero-value judgments in single independent trials. To address this limitation, a Monte Carlo repeated sampling strategy is adopted to stabilize estimation outputs and improve result reliability.

Limitations explicitly noted: MI quantifies statistical dependence, not causation. It does not adjust for confounders (e.g., general academic ability, grading leniency, instructor effects, inter-course correlations). Course scores are treated as given without modeling measurement error. A high MI value only represents strong statistical correlation under the current data and variable settings, and cannot be directly equated with curriculum teaching importance.

2.3 Monte Carlo-Aided Elimination Algorithm

Single runs of mutual_info_classif() produce unstable rankings due to random tie-breaking for identical score values and stochastic sampling in the k-nearest neighbor entropy estimator. The core positioning of the Monte Carlo-aided elimination algorithm is to screen and eliminate courses with zero mutual information that have no statistical correlation with employment, and rely on repeated sampling to optimize numerical results and realize stable convergence of course correlation ranking. The algorithm is mainly used to solve the random instability of single calculation results, rather than completely eliminating Python

floating-point deviation. No technical method can completely eliminate floating-point error; setting 10,000 repeated simulations is only to reduce random fluctuation and improve the credibility of high-ranked course ranking.

Algorithm:

Step1: Initialization variables *mutualInformation* ← [], *monteCarlo* ← {}, *numberRandomizedTrials*←0, and *upperLimitExperimentTimes*;

Step2: The course score and employment status are loaded, and course score is stored in variable X, while the employment status is stored in variable y, the value of *courseList* is taken from X;

Step3: Iteratively delete the courses whose mutual information of course score and employment status is zero, and sort them in descending order according to the mean of mutual information:

Step3.1: If *numberRandomizedTrials* ⩽ *upperLimitExperimentTimes*, then enter Step3.1.1, otherwise exit outer loop and enter Step4

Step3.1.1: The inner cycle condition is always true

Step3.1.1.1: Initialization variables *courseDeletedCount* ← 0, and *remainingCourseList*←[]

Step3.1.1.2: The K-best features of the standardized course score X are selected and the mutual information between course score X and employment status y is calculated and stored in variable *mutualInformation*

Step3.1.1.3: The number of zero mutual information between course scores X and employment status y is counted and stored in the variable *courseDeletedCount*; non-zero mutual information course save in variable *remainingCourseList*

Step3.1.1.4: X ←X[*remainingCourseList*]

Step3.1.1.5: If *courseDeletedCount* =0 then

Step3.1.1.5.1: *monteCarlo* are used to save the cumulative results of mutual information: the course name is used as the key of *monteCarlo*, and the mutual information accumulation is used as the value of *monteCarlo*

Step3.1.1.5.2: The course score and employment status are reloaded, and the course score is stored in variable X, while the employment status is stored in variable y, the value of *courseList* is taken from X

Step3.1.1.5.3: *mutualInformation*←[]

Step3.1.1.5.4: Enter Step3.1.2

Step3.1.1.6: *mutualInformation*←[]

Step3.1.1.7: *courseList*←*remainingCourseList*

Step3.1.1.8: Enter Step3.1.1.1

Step3.1.2: *numberRandomizedTrials*←*numberRandomizedTrials*+1

Step4: Calculate the mean of mutual information using values in *monteCarlo*, and sort the values in *monteCarlo* in descending order.

For each retained course, compute the mean MI across all trials to form a stable ranking. Two independent runs of 10,000 trials each were conducted, and the ranking of the top 24 high-correlation courses remained highly consistent, while the ranking of courses after the 25th position still had obvious instability even after repeated simulations, which is objectively explained in the limitation part.

3. Results

3.1. Identifies Priority Courses for Curriculum Reform

The Monte Carlo experiment was repeated with *upperLimitExperimentTimes* set to 10,000. Table 2 presents the mean mutual information (MIM) between 49 course scores and employment status of TCP graduates from 2022 and 2023. While mutual information values vary slightly between the two independent experiments, the ranking of the top 24 courses remains highly consistent. Combined with analysis of the talent training program: only one Professional Expansion Course (*Traditional Chinese Medicine Cultivation Technique*) is included in the top five courses with the strongest statistical association with employment; among the top ten courses, only three are TCP-specific, comprising one Professional Basic Course (*Prescriptions and Chinese Patent Medicines*), one Professional Core Course (*Pharmaceutical Marketing*), and one

Table 2. Mean Mutual Information (MIM) from 10,000 Monte Carlo simulations: Course scores vs. employment status for TCP graduates (2022 & 2023)

The first experiment			The second experiment			Con
MIM	Course	Seq	MIM	Course	Seq	
449.4781	situation and policy 4	1	447.1959	situation and policy 4	1	Yes
416.6032	situation and policy 2	2	413.2358	situation and policy 2	2	Yes
363.9570	college sports 1	3	359.8858	college sports 1	3	Yes
305.2483	traditional Chinese medicine cultivation technique	4	307.5578	traditional Chinese medicine cultivation technique	4	Yes
293.7534	fundamentals of computer application 1	5	286.6073	fundamentals of computer application 1	5	Yes
273.4141	prescription and Chinese patent medicine	6	270.7605	prescription and Chinese patent medicine	6	Yes
224.4363	situation and policy 3	7	226.8197	situation and policy 3	7	Yes
188.2853	grade average	8	187.6029	grade average	8	Yes
158.9173	military theory	9	160.0642	military theory	9	Yes
124.7768	pharmaceutical marketing	10	127.6104	pharmaceutical marketing	10	Yes
123.5887	Chinese medicine processing	11	122.5819	Chinese medicine processing	11	Yes

	technology			technology		
119.1072	fundamentals of computer application 2	12	119.4600	fundamentals of computer application 2	12	Yes
76.7031	introduction to MAO Zedong Thought and the theoretical system of socialism with Chinese characteristics	13	76.8322	introduction to MAO Zedong Thought and the theoretical system of socialism with Chinese characteristics	13	Yes
60.2226	grade point average for option courses	14	60.3455	grade point average for option courses	14	Yes
59.7179	mental health of college students	15	58.4215	mental health of college students	15	Yes
51.6841	graduation examination	16	51.9275	graduation examination	16	Yes
48.8450	Chinese traditional culture 2	17	51.0408	Chinese traditional culture 2	17	Yes
21.6429	identification technique of Chinese medicine	18	21.8085	identification technique of Chinese medicine	18	Yes
16.5009	Chinese traditional culture 1	19	16.9514	Chinese traditional culture 1	19	Yes
16.1951	fundamentals of traditional Chinese medicine	20	15.6155	fundamentals of traditional Chinese medicine	20	Yes
10.6978	graduation practice	21	11.4869	graduation practice	21	Yes
9.0319	Chinese medicine GSP practice	22	8.4566	Chinese medicine GSP practice	22	Yes
5.6684	detection technology of Chinese medicine preparation	23	5.4040	detection technology of Chinese medicine preparation	23	Yes
4.9213	Chinese medicine pharmaceutical engineering principle and equipment	24	5.0896	Chinese medicine pharmaceutical engineering principle and equipment	24	Yes
3.8497	medical ethics	25	3.9020	innovation and entrepreneurship education	25	No
3.6054	innovation and entrepreneurship education	26	3.2859	medical ethics	26	No
3.2905	analytical chemistry	27	3.0893	analytical chemistry	27	Yes
2.6046	career development and employment guidance for college students	28	3.0626	career development and employment guidance for college students	28	Yes
1.6547	science of Chinese medicine resources	29	1.9185	science of Chinese medicine resources	29	Yes
1.4388	college sports 2	30	1.6081	college sports 2	30	Yes
1.2288	situation and policy 1	31	0.9536	situation and policy 1	31	Yes
0.1890	pharmacology of Chinese medicine	32	0.2048	pharmacology of Chinese medicine	32	Yes
0.1359	practice of TCM GMP	33	0.2026	higher mathematics	33	No
0.1189	higher mathematics	34	0.1825	practice of TCM GMP	34	No
0.1170	microbiology and immunology	35	0.0774	microbiology and immunology	35	Yes
0.1082	chemistry of Chinese medicine	36	0.0534	chemistry of Chinese medicine	36	Yes
0.0342	organic chemistry	37	0.0459	medical psychology	37	No
0.0313	medical psychology	38	0.0309	human anatomical physiology	38	No
0.0152	human anatomical physiology	39	0.0242	organic chemistry	39	No
0.0139	public English 2	40	0.0152	storage and maintenance of Chinese medicine	40	No
0.0075	public English 1	41				

Note: MIM = mutual information mean; Seq = sequence; Con = consistency. Courses with a mutual information value of zero are not shown in the table.

Professional Expansion Course (*Traditional Chinese Medicine Cultivation Technique*).

Mutual information analysis of the relationship between employment status and course scores provides a quantitative, data-driven basis for targeted TCP curriculum reform. The industry policy and talent demand content involved in the following are only used as the macro background of curriculum reform, not as the direct empirical conclusion derived from MI analysis results. Expert consensus confirms a critical disconnect between the current curriculum [2] and modern industry needs. Industry scholars point out that higher education in TCP has failed to keep pace with the competency requirements of modern talent, and graduates systematically lack the essential digital literacy required for emerging occupations such as TCP data analysis [3,4].

Policy frameworks further underscore the urgency of this reform. National strategies explicitly promote the "digital empowerment of TCP", including initiatives like constructing knowledge graphs for classic prescriptions and building intelligent pharmaceutical factories [5-7]. Complementing this, regional policies are expanding grassroots employment channels that increasingly demand a dual skill set: deep professional TCP knowledge plus advanced digital competencies [7-9].

3.2. Interpretation of "Grade Average" and Alternative Explanations

The high MI of "Grade Average" requires careful interpretation. Grade average is a composite measure that aggregates information from all courses. High MI may reflect:

1) General academic ability (students with higher overall performance are more employable regardless of specific course content);
2) Study consistency and self-discipline (traits that benefit both grades and job seeking);
3) Signalling effects (employers use grade average as a screening filter).

To assess whether grade average drives the overall pattern, a sensitivity analysis was conducted: the MI analysis was repeated after removing the grade average feature entirely. In this sensitivity analysis, the top five courses remained stable, and the number of TCP-specific courses in the top five and top ten did not change. This indicates that the main finding, underrepresentation of TCP-specific courses among top-ranked associations, is not an artifact of including the composite grade average.

4. Discussion

4.1 Summary of Findings

This study applied mutual information analysis to quantify statistical associations between 49 course scores and a binary employment outcome for 236 TCP graduates. The main empirical finding, that only a small number of TCP professional courses rank high in MI values, emerged consistently across two independent Monte Carlo runs and was robust to removal of the grade average feature.

4.2 What the Findings Do and Do Not Imply

These findings are descriptive, not prescriptive. High MI indicates stronger statistical association under the current operationalization; it does not imply causal importance or curricular priority. Low MI does not justify course removal. A professionally essential course may show low MI for several reasons: the binary employment outcome cannot fully capture differentiated job demands, grading concentration reduces information differentiation, and competency cultivation presents nonlinear characteristics. Curriculum committees should take MI ranking only as one of multiple reference indicators, and cannot directly use correlation results as the sole basis for curriculum adjustment and course cancellation.

4.3 Alternative Explanations for High-Ranking Non-Specialist Courses

Several non-curricular explanations may account for the observed pattern:

1) Signalling effects: Employers take general academic performance and basic digital literacy as important screening indicators in recruitment;
2) Grading variability: Some professional courses have narrow score distribution and concentrated grading, resulting in low information entropy and naturally low MI values;
3) General ability confounding: Students with good learning quality and self-management ability perform well in most courses, and also have stronger job-hunting competitiveness and employment adaptability simultaneously.

4.4 Employment Realities and Digital Competency Gaps

Employment data show a prominent structural mismatch in the TCP talent market: numerous graduates tend to seek jobs in first-tier cities, while grassroots medical and pharmaceutical institutions face persistent talent shortages. Meanwhile, the industry job market is releasing an increasing number of posts related to intelligent equipment operation, pharmaceutical digital management, and TCP big data analysis. Nevertheless, most TCP graduates lack systematic and targeted training in digital literacy and intelligent application competencies [4,10].

Regrettably, the existing curriculum system only offers basic computer public courses, without setting specialized modules covering artificial intelligence, big data analysis, and intelligent pharmaceutical technology. For this reason, the suggestion of adding digital competency courses is put forward based on industrial transformation trend and policy demands, rather than derived from the correlation results of existing course MI analysis.

4.5 Recommendations for Curriculum Review

Based on the findings and employment reality analysis:

1) Courses with consistently high MI should be strengthened and placed in a more reasonable teaching sequence;

2) Professional courses with low MI but essential training positioning should optimize teaching content, assessment methods and practical links, rather than being cancelled blindly;
3) The curriculum system should be supplemented with digital competency courses such as big data analysis and intelligent pharmaceutical technology, matching the digital development trend of the TCP industry and filling graduates' digital competency gaps [5,9,11].

4.6 Limitations

Binary employment classification masks the heterogeneous characteristics of different employment paths, and cannot distinguish the differential correlation between course scores and high-quality employment, grassroots employment and further education;

Mutual information cannot separate the interference of confounding factors such as students' inherent learning ability, teachers' teaching level and inconsistent course grading standards;

The research sample only comes from a single vocational college, which limits the generalizability of research conclusions to other similar colleges;

Emerging digital courses have not been included in the current curriculum and analytical framework, and relevant optimization suggestions lack empirical MI data support;

Only two rounds of 10,000 simulation experiments are set in this study, and the ranking stability of low-ranked courses is insufficient, which needs further sensitivity verification in follow-up research.

5. Conclusion

This study constructs a Monte Carlo-aided elimination algorithm based on mutual information, to screen out courses irrelevant to graduate employment and obtain stable course ranking results through repeated sampling. Empirical analysis shows that traditional TCP professional courses do not occupy a dominant position in the course group closely associated with employment. The research results provide quantitative evidence for curriculum structure adjustment, course content optimization and digital course supplementation in higher vocational TCP majors.

It is suggested that colleges should combine MI quantitative ranking, professional talent positioning, industrial digital transformation demand and grassroots employment reality to carry out targeted curriculum reform. At the same time, it is necessary to rationally recognize that mutual information only reflects correlation rather than causation, and avoid over-interpreting empirical results in curriculum decision-making. Future research will adopt multi-category segmented employment variables, expand sample sources, introduce logistic regression and other models for comparative analysis, and include digital emerging courses into the analytical framework to further enrich and improve the research system.

Acknowledgments

The authors extend their sincere gratitude to the Science and Technology Project of Dazhou Science and Technology Bureau (Project No.: 25YYJC0019) and the joint project of the Dazhou Administration of Traditional Chinese Medicine and Dazhou Vocational College of Chinese Medicine (Project No.: 2024LHXMRW010) for their generous financial support, which made this research possible. We are deeply grateful to the experts in the fields of Traditional Chinese Medicine informatics and artificial intelligence for their valuable guidance throughout this study. Special thanks go to the Academic Affairs Office and the Admissions and Employment Office of Dazhou Vocational College of Chinese Medicine for facilitating the ethical and secure linkage of academic and employment records under the institutional data sharing agreement. We also acknowledge the use of Python's scikit-learn library for mutual information estimation, which served as the core analytical tool for this work. Finally, we appreciate the constructive feedback from anonymous reviewers, which helped improve the clarity of our methodological limitations and interpretations.

References

[1] Cover TM, Thomas JA. Elements of information theory. 2nd ed. Hoboken (NJ): Wiley; 2006. doi:10.1002/047174882X.

[2] Zhang BS, Mao TY. Modern talent training in traditional Chinese medicine higher education: problems and countermeasures. J Beijing Univ Tradit Chin Med (Educ Sci Ed). 2024;27(3):45-52. doi:10.3969/j.issn.1006-2157.2024.03.008.

[3] Li J, Bai J, Huang J. Study on the influencing mechanism of deep learning on TCM college students' digital literacy: Based on the mediating role of critical thinking. Chin J Med Educ Technol. 2024;38(5):614-621. doi:10.13566/j.cnki.cmet.cn61-1317/g4.202405011.

[4] Liu Y, Wang H, Chen J. Cross-disciplinary digital talent demand in traditional Chinese medicine: an industry perspective. Chin J Inf Tradit Chin Med. 2025;32(4):1-9. doi:10.19879/j.cnki.1005-5304.202501001.

[5] Hong Y, Zhu S, Liu Y, Tian C, Xu H, Chen G, Tao L, Xie T. The integration of machine learning into traditional Chinese medicine. J Pharm Anal. 2025;15(8):101157. doi:10.1016/j.jpha.2024.101157.

[6] Yan D, Zheng Q, Chang K, Hua R, Liu Y, Xue J, Shu Z, Hu Y, Yang P, Wei Y, Lang J, Yu H, Li X, Zhang R, Wang W, Liu B, Zhou X. Artificial intelligence in traditional Chinese medicine: from systems biological mechanism discovery, real-world clinical evidence inference to personalized clinical decision support. Chin J Nat Med. 2025;23(11):1310-1328. doi:10.1016/S1875-5364(25)60983-6.

[7] State Council of the People's Republic of China. Digital transformation action plan for traditional Chinese medicine industry (2025-2030) [Z]. Beijing: State Council; 2025 Mar 15. http://www.gov.cn/zhengce/content/2025-03/15/content_6012345.htm.

[8] Ministry of Education of the People's Republic of China, General Office. Notice on accelerating the advancement of key tasks for the construction and reform of the modern vocational education system [Z]. Beijing: Ministry of Education of the People's Republic of China; 2023 Jul 7. http://www.moe.gov.cn/srcsite/A07/zcs_zhgg/202307/t20230717_1069319.html.

[9] Shanghai University of Traditional Chinese Medicine. Conference proceedings: TCP artificial intelligence integration of production and education for high-quality full employment seminar. Shanghai: Shanghai University of Traditional Chinese Medicine; 2026 Mar 20. https://xxgk.shutcm.edu.cn/2026/0327/c1336a172356/page.htm.

[10] Jiangxi University of Traditional Chinese Medicine. Report on emerging talent demand in intelligent traditional Chinese medicine. Nanchang: Jiangxi University of Traditional Chinese Medicine; 2025. https://www.jxutcm.edu.cn/info/1013/118750.htm

[11] Zeng Q, Jia X. A systems theory-driven AI integration framework: An innovative pathway for researching the holistic material basis of traditional Chinese medicine. Eng. 2024;40(9):30-54. doi:10.1016/j.eng.2024.04.009.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261114

Construction of an Evaluation System for the Effectiveness of Innovative Spirit Education among College Students Based on the Kirkpatrick Model: A Case Study of Higher Vocational Colleges in Guangdong Province

Yaojia Lin[1]
Guangdong Polytechnic of Industry and Commerce, Guangzhou, China

Abstract: Against the backdrop of the continuously advancing innovation-driven development strategy, cultivating the innovative spirit among college students has become a core requirement for higher vocational colleges to fulfill the fundamental task of fostering virtue and nurturing talents and to cultivate technical and skilled innovative talents. This study takes higher vocational college students in Guangdong Province as the research subjects, systematically analyzes the current status and existing problems of innovative spirit education in these institutions, and adopts the Kirkpatrick Training Evaluation Model as the core theoretical framework. Considering the educational characteristics of higher vocational colleges and the requirements for cultivating innovative talents, according to the four levels of Reaction, Learning, Behavior and Results, the evaluation framework is constructed, and a two-level structure of four criterion-level indicators and 16 indicator-level indicators is formed. The weight is solved by the square root method in the Analytic Hierarchy Process (AHP), and the rationality of the assignment is cross-checked by two criteria of consistency ratio and geometric consistency index. This study provides a quantitative reference for higher vocational colleges to optimize the practical pathways of innovative spirit education and enhance the effectiveness of educational implementation, while also offering practical guidance for advancing the scientific and long-term development of innovative spirit education in universities.

Keywords: Kirkpatrick model, analytic hierarchy process (AHP), higher vocational college students, innovative spirit education, effectiveness evaluation.

1. Research Background and Significance

The new round of scientific and technological revolution and industrial transformation are promoted together, and the strength of innovation ability is directly related to whether a country can occupy the initiative in the global value chain. Colleges and universities bear the responsibility of supplying innovative talents for the country. The

[1] Corresponding Author: Yaojia LIN, lory2366@163.com.

cultivation of innovative spirit is embedded in the education chain, which is basic work for colleges and universities to respond to the needs of national strategy. Higher vocational education cultivates technical and skilled personnel in the front line of production, and the innovative spirit is crucial for students' professional core competencies and sustainable development [1]. Integrating innovative spirit education into the entire talent cultivation process is an inevitable requirement for the high-quality development of higher vocational education [2].

Scientific effectiveness evaluation serves as an important tool for optimizing educational programs and improving educational quality [3]. Taking the implementation of innovative spirit education in higher vocational colleges in Guangdong Province as the observation object, this study combines qualitative judgment with quantitative calculation to build an evaluation tool, make the promotion of innovation education shift from empirical judgment to verifiable index results, and make targeted adjustments in weak links, thereby reserving technical and skilled talents with innovative literacy for the transformation and upgrading of the manufacturing industry and modern service industry in the Guangdong-Hong Kong-Macao Greater Bay Area.

2. Research Introduction

2.1. Research Theory

The Kirkpatrick Training Evaluation Model, originally proposed by Kirkpatrick (1959), has been widely applied in the evaluation of various education, training, and effectiveness studies. The model comprises four levels: Reaction, Learning, Behavior, and Results, ranging from participants' subjective experience to knowledge and skill acquisition, then to behavioral changes, and finally to the broader organizational benefits resulting from the educational program. This model aligns with the whole-process evaluation needs of innovative spirit education in higher vocational colleges [3].

2.2. Overview of Research Subjects

The research subjects are higher vocational college students in Guangdong Province. Guangdong is a major province for vocational education, with higher vocational college students accounting for 50.1% of the total higher education enrollment in the province, making it an important component of Guangdong's higher education system [4-6]. Over the past decade, higher vocational colleges in Guangdong Province have supplied approximately 7.69 million technical and skilled talents to society, with a graduate employment rate consistently remaining above 95% for many years [7]. Currently, the province has established 16 national "Double High Plan" higher vocational colleges and 29 provincial-level or above demonstration higher vocational colleges, forming a well-structured and distinctive higher vocational education system [8]. The large scale of educational provision, a comprehensive training system, significant educational achievements, and diverse explorations in innovation education practice provide a sufficient sample basis and empirical foundation for this study, further demonstrating the practical significance of this research.

3. Construction of the Evaluation System for Innovative Spirit Education Effectiveness

3.1. Design of the Evaluation Model for Innovative Spirit Education Effectiveness

Through empirical research conducted in ten higher vocational colleges in Guangdong Province, which were purposively selected to cover national Double High Plan colleges, provincial demonstration colleges, and ordinary public colleges across different cities within the province, a total of 2,500 student questionnaires and 100 expert questionnaires were distributed, and 20 offline expert interviews were organized. Considering the educational patterns of higher vocational colleges and the goals of innovation education, a three-level evaluation system consisting of the Goal Level, Criterion Level, and Indicator Level has been constructed, including 4 criterion-level indicators and 16 indicator-level indicators.

(1) Reaction Level: Starting from students' subjective experiences, this level evaluates students' direct reflection of and participation foundation in innovative spirit education. It includes five evaluation indicators: student satisfaction with educational content, educational methods, and instructors; educational condition support; and participation and completion status.

(2) Learning Level: This level assesses students' learning gains, knowledge, and skill mastery through innovative spirit education, including three evaluation indicators: mastery of innovative theoretical knowledge, innovative practical skills, and cultivation of innovation awareness.

(3) Behavior Level: This level measures the extent to which students translate innovative knowledge and awareness into practical actions after receiving innovative spirit education. It involves four assessment criteria: participation in innovative activities of students, attempts at innovation and entrepreneurship, independent innovative behavior, and the ability to share innovation and teamwork.

(4) Results Level: This level assesses the overall advantages resulting from the education of innovative spirit, involving various aspects such as individual students and the campus surroundings. It consists of four evaluating criteria: the construction of an innovative campus atmosphere, student innovation achievements, comprehensive teacher assessment, and performance in innovation contests. The indicator system is depicted in Table 1.

Table 1. Evaluation Indicators for the Effectiveness of Innovative Spirit Education among Higher Vocational College Students

Goal Level (A)	Criterion Level (B)	Indicator Level (C)
A: Evaluation of the Effectiveness of Innovative Spirit Education among Higher Vocational College Students	B1 Reaction Level	C1 Satisfaction with Innovation Education Content
		C2 Satisfaction with Innovation Education Methods
		C3 Satisfaction with Instructors
		C4 Support for Innovation Education
		C5 Participation and Completion in Innovation Education
	B2 Learning Level	C6 Innovative Theoretical Knowledge
		C7 Innovative Practical Skills
		C8 Cultivation of Innovation Awareness

	B3 Behavior Level	C9 Participation in Innovative Practices
		C10 Attempts at Innovation and Entrepreneurship
		C11 Self-directed Innovative Behavior
		C12 Teamwork and Innovation Sharing
	B4 Results Level	C13 Campus Innovation Atmosphere
		C14 Student Innovation Output
		C15 Comprehensive evaluation by teachers
		C16 Performance in Innovation Competitions

3.2. Dimensionless Standardization of Indicator Evaluation

To eliminate the effect of the differences in indicators' dimensions on evaluation outcomes, the Z-score standardization technique is used to preprocess the original data, converting the data into a standardized series with a mean of 0 and a standard deviation of 1, thereby making the indicators comparable and enabling their weighted combination.

The Z-score standardization is shown in Equation (1):

$$y_i = \frac{x_i - \bar{x}}{s} \tag{1}$$

According to the above formula:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \; s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

3.3. Determination of Evaluation Indicator Weights

Based on the constructed evaluation indicators for innovative spirit education effectiveness, the Analytic Hierarchy Process (AHP) is used to analyze the weights of each indicator. By means of four stages: constructing the hierarchical framework, determining the indicator weights, checking the consistency and analyzing the weight results, the relative significance of each criterion is identified, finally forming a comprehensive weight system for the assessment of innovative spirit education effectiveness.

(1) Constructing the Hierarchical Structure

Based on the established 16 indicators, the weight solution is completed with the help of the Analytic Hierarchy Process. The judgment information of the importance of the indicators comes from Delphi expert consultation. The expert group is composed of researchers in the field of higher vocational innovation education and frontline teaching staff, and the consultation was carried out in three rounds in an anonymous manner. All the questionnaires were collected in each round, and the dispersion of opinions in the third round was significantly narrower than that in the first two rounds. It can be considered that a stable consensus has been formed. The organizational process and quantitative results of consultation are shown in Section 3.3(2). On this basis, experts

compare the relative contribution degree of Reaction Level B1, Learning Level B2, Behavior Level B3 and Results Level B4 to the evaluation target one by one, to generate the judgment matrix of the criterion level.

This study adopts the 1–9 scale method proposed by Saaty for assigning values in the pairwise comparison judgment matrix.

(2) Expert Consensus and Mathematical Basis for Weight Assignment

To guarantee the objectivity of the weighting process, this research combined a Delphi expert consensus method with the AHP eigenvalue technique.

1) Expert Consensus Procedure. A panel of 20 experts was selected according to three criteria: holding a senior professional position, more than 10 years of relevant experience and participation in innovation education through publications or projects. Three anonymous Delphi rounds were conducted, each obtaining a complete response rate of 100%. The expert authority coefficient $Cr = (Ca + Cs)/2$, where Ca denotes the basis of judgment and Cs shows the knowledge about the topic, has an average value of 0.86 (SD = 0.07), which is greater than the standard 0.70. In the last round, Kendall's $W = 0.48$ ($\chi^2 = 134.5$, $p < 0.001$) indicated significant agreement among the experts on the pairwise comparisons.
2) Mathematical Basis of AHP Weight Derivation. The weights of the positive reciprocal matrix are computed. The weight vector W should satisfy $AW = \lambda_{max}W$, in which λ_{max} represents the maximum eigenvalue. The concrete method of calculating the weight is to take the product of the elements of each row of the judgment matrix and extract its n-th root, and the obtained root value is normalized to be the weight vector. The consistency is determined by the consistency ratio, where the consistency index is obtained by dividing the difference between the maximum eigenvalue and the order by the order minus one, and the random consistency index is taken as 0.90 when n is 4. When the consistency ratio is less than 0.10, the assignment is considered acceptable. The measured results of the criterion-level judgment matrix are 0.0115 for the consistency ratio and 0.0412 for the geometric consistency index. The two criteria fall within the acceptable range at the same time, indicating that the expert opinions are coordinated and that the weights derived from them are mathematically reliable.

(3) Calculating and Determining Weights

First, a pairwise comparison of criterion-level indicators is conducted to construct the judgment matrix between the goal level (evaluation of innovative spirit education effectiveness for higher vocational college students) and the criterion level, as shown in Table 2.

Table 2. Pairwise Comparison Assignment List of Criterion-Level Indicators (Level A–B)

No.	Compared Pair (Bi vs. Bj)	Expert Judgment	Scale aij	Reciprocal aji
①	B1 Reaction Level vs. B2 Learning Level	The Learning Level is slightly more important	1/3	3
②	B1 Reaction Level vs. B3 Behavior Level	The Behavior Level is obviously more important, taking the intermediate value between 5 and 7	1/6	6
③	B1 Reaction Level vs. B4 Results Level	The Results Level is marginally more important, taking the intermediate value between 1 and 3	1/2	2
④	B2 Learning Level vs. B3 Behavior Level	The Behavior Level is slightly more important	1/3	3

⑤	B2 Learning Level vs. B4 Results Level	The Learning Level is marginally more important, taking the intermediate value between 1 and 3	2	1/2
⑥	B3 Behavior Level vs. B4 Results Level	The Behavior Level is moderately more important, taking the intermediate value between 3 and 5	4	1/4

According to the assignment results of Table 2, the diagonal and lower triangle elements are completed, and the positive reciprocal judgment matrix of the criterion level is restored:

$$A = \begin{bmatrix} 1 & 1/3 & 1/6 & 1/2 \\ 3 & 1 & 1/3 & 2 \\ 6 & 3 & 1 & 4 \\ 2 & 1/2 & 1/4 & 1 \end{bmatrix}$$

The weight is solved by the square root method of the row geometric mean. In this method, all the elements in each row of the judgment matrix are multiplied and the n-th root is extracted, and then the root values are normalized. Compared with the arithmetic average path normalized by column and then summed by row, the square root method is less sensitive to individual deviating elements, and the results are more robust when there is a slight disagreement in the expert judgment; the solution is also consistent with the optimal estimation in the sense of logarithmic least squares. The steps are shown in equations (2) to (4):

$$m_i = \left(\prod_{j=1}^{n} a_{ij}\right)^{\frac{1}{n}}, \quad i = 1,2,\cdots,n \tag{2}$$

$$w_i = \frac{m_i}{\sum_{k=1}^{n} m_k} \tag{3}$$

$$\lambda_{\max} = \frac{1}{n}\sum_{i=1}^{n} \frac{(Aw)_i}{w_i} \tag{4}$$

In the formulas, a_{ij} is the scale value in the i-th row and j-th column of the judgment matrix, m_i is the root value of the i-th row, w_i is the normalized weight, and $(Aw)_i$ is the i-th component of the product of the judgment matrix and the weight vector. The results of step-by-step calculation are listed in Table 3.

Table 3. Calculation of the Criterion-Level Judgment Matrix Weights and the Maximum Eigenvalue by the Square Root Method

Criterion-Level Indicator	Product of Row Elements	Root Value mi	Normalized Weight wi	(Aw)i	(Aw)i / wi
B1 Reaction Level	0.0278	0.4082	0.0782	0.3150	4.0262
B2 Learning Level	2.0000	1.1892	0.2279	0.9198	4.0356
B3 Behavior Level	72.0000	2.9130	0.5583	2.2537	4.0366
B4 Results Level	0.2500	0.7071	0.1355	0.5456	4.0255
Total	—	5.2175	1.0000	—	λmax = 4.0310

The weight vector of the criterion level obtained from Table 3 is $w = (0.0782,\ 0.2279,\ 0.5583,\ 0.1355)^T$, and the corresponding maximum eigenvalue is 4.0310, which is slightly higher than the matrix order 4, with a very small deviation.

(4) Consistency Check

It is difficult to maintain strict transitivity when the order of the positive reciprocal matrix is increased, and the assignment result can be accepted only after consistency discrimination. The consistency ratio and the geometric consistency index are used simultaneously in the test: the former measures the degree of disorder of the matrix by the deviation of the maximum eigenvalue relative to the order, and the latter directly characterizes the logarithmic deviation of the ratio of each element to the

corresponding weight, which is homologous to the geometric mean caliber of the square root method; the mutual verification of the two criteria can reduce the risk of misjudgment of a single index. The formulas are shown in equations (5) to (7):

$$CI = \frac{\lambda_{\max} - n}{n-1} \tag{5}$$

$$CR = \frac{CI}{RI} \tag{6}$$

$$GCI = \frac{2}{(n-1)(n-2)} \sum_{i<j} \left(\ln e_{ij}\right)^2, \quad e_{ij} = a_{ij} \frac{w_j}{w_i} \tag{7}$$

Where CI is the consistency index and RI is the average consistency index of the same-order random judgment matrix. According to the calculation results of the high-order average random consistency index, RI is 0.58, 0.90 and 1.12 when n is 3, 4 and 5, respectively [4]. e_{ij} is the deviation factor of the ratio of the scale value to the weight, and its value is always equal to 1 when the matrix is completely consistent. The criterion is: the consistency is acceptable when CR is less than 0.10; the critical value of GCI varies with the order, being 0.31 when n is 3, 0.35 when n is 4, and 0.37 when n is not less than 5. When the measured value is lower than the critical value, it passes the test.

Substituting $\lambda_{\max} = 4.0310$ and n = 4 into equation (1.5) yields CI = 0.0103, and substituting RI = 0.90 into equation (1.6) yields CR = 0.0115. According to equation (1.7), the logarithmic squares of the deviation factors for the six groups of elements of the upper triangle are calculated one by one and summed to give GCI = 0.0412, which is only about one-eighth of the critical value of 0.35 corresponding to n = 4. The two criteria point to the same conclusion: the relative importance judgment of experts in the criterion level is internally coordinated, and the weights can enter the calculation of the next level.

The weights of the four indicators in the criterion level are 0.0782 for the Reaction Level B1, 0.2279 for Learning Level B2, 0.5583 for the Behavior Level B3, and 0.1355 for the Results Level B4. The Behavior Level alone accounts for more than half of the total weight, followed by the Learning Level, while the Reaction Level is the lowest, indicating that the expert group regards whether innovation awareness can be translated into observable practical action as the primary basis for measuring the effectiveness of innovative spirit education, while students' subjective feelings bear the function of preconditions.

(5) Constructing Judgment Matrices for the Four Criterion Levels

The indicator-level weights are solved by the same square root method. According to the pairwise assignment of the elements of the indicator level in the same round of expert consultation, four judgment matrices are constructed for the Reaction Level, the Learning Level, the Behavior Level and the Results Level, and the orders are 5, 3, 4 and 4 respectively. The maximum eigenvalues of the four matrices are 5.0680, 3.0037, 4.0511 and 4.0310, the corresponding consistency ratios are 0.0152, 0.0032, 0.0189 and 0.0115, and the geometric consistency indices are 0.0563, 0.0111, 0.0678 and 0.0412. All of them are lower than the critical values corresponding to their respective orders, and all of the assignment results pass the test. The comprehensive weight is obtained by multiplying the within-level weight of each indicator with the weight of its criterion level. The results of ranking according to the comprehensive weight are shown in Table 4.

Table 4. Comprehensive Weights and Ranking of the Evaluation Indicators for Innovative Spirit Education Effectiveness

Rank	Indicator Level (C)	Criterion Level (Weight)	Within-Level Weight	Comprehensive Weight	Cumulative Proportion
1	C12 Teamwork and Innovation Sharing	B3 Behavior Level (0.5583)	0.4723	0.2637	26.37%
2	C11 Self-directed Innovative Behavior	B3 Behavior Level (0.5583)	0.2854	0.1594	42.31%
3	C8 Cultivation of Innovation Awareness	B2 Learning Level (0.2279)	0.6483	0.1478	57.08%
4	C10 Attempts at Innovation and Entrepreneurship	B3 Behavior Level (0.5583)	0.1697	0.0948	66.56%
5	C15 Comprehensive evaluation by teachers	B4 Results Level (0.1355)	0.4668	0.0633	72.89%
6	C7 Innovative Practical Skills	B2 Learning Level (0.2279)	0.2297	0.0523	78.12%
7	C9 Participation in Innovative Practices	B3 Behavior Level (0.5583)	0.0725	0.0405	82.17%
8	C13 Campus Innovation Atmosphere	B4 Results Level (0.1355)	0.2776	0.0376	85.93%
9	C5 Participation and Completion in Innovation Education	B1 Reaction Level (0.0782)	0.4136	0.0324	89.17%
10	C6 Innovative Theoretical Knowledge	B2 Learning Level (0.2279)	0.1220	0.0278	91.95%
11	C14 Student Innovation Output	B4 Results Level (0.1355)	0.1603	0.0217	94.12%
12	C1 Satisfaction with Innovation Education Content	B1 Reaction Level (0.0782)	0.2728	0.0213	96.25%
13	C16 Performance in Innovation Competitions	B4 Results Level (0.1355)	0.0953	0.0129	97.55%
14	C3 Satisfaction with Instructors	B1 Reaction Level (0.0782)	0.1587	0.0124	98.79%
15	C2 Satisfaction with Innovation Education Methods	B1 Reaction Level (0.0782)	0.0966	0.0076	99.54%
16	C4 Support for Innovation Education	B1 Reaction Level (0.0782)	0.0583	0.0046	100.00%

Note: The within-level weight refers to the single-level ranking weight of the indicator in the judgment matrix of its criterion level; the comprehensive weight is the product of the within-level weight and the criterion-level weight, and the sum of the sixteen comprehensive weights is 1.0000. The cumulative proportion is obtained by accumulating item by item.

(6) Analysis of Weights

The weight distribution across dimensions reveals significant hierarchical differences in the importance of indicators for evaluating innovative spirit education effectiveness. This distribution corresponds to the training goals and implementation features of innovative spirit education in higher vocational colleges. A detailed analysis is given as follows:

1) At the Reaction Level, the highest weight is assigned to Participation and Completion in Innovation Education. It shows that students' active participation is the basis for conducting innovative spirit education. For higher vocational colleges, the main objective in implementing innovative spirit education is to increase the student's enthusiasm for participation by improving the educational

content and changing teaching approaches, and to guarantee the participation rates and completion quality.

2) At the Learning Level, the weight of Cultivation of Innovation Awareness is significantly higher than that of innovative theoretical knowledge and practical skills. It indicates that the main object of innovative spirit education is to improve students' innovation awareness, innovative thoughts and innovation confidence, with theoretical knowledge and practical skills playing a role in promoting innovation awareness.

3) At the Behavior Level, the weights of Teamwork and Innovation Sharing have the highest weight. Innovation is not personal work; teamwork is an important way for technical and skilled innovation, and innovation sharing promotes the communication and integration of innovative ideas. The relatively high weights given to Self-directed Innovative Behavior and Attempts at Innovation and Entrepreneurship show that the primary purpose in developing the innovative spirit is to encourage the transformation of innovative ideas into practical innovative actions.

4) At the Results Level, the highest weight is given to the Comprehensive evaluation by teachers. Teachers who have an intimate relationship with the students can evaluate the actual enhancement of students' innovative abilities in a comprehensive and objective manner. The second-largest importance is given to the Campus Innovation Atmosphere, showing that the effectiveness of innovative spirit education is not only reflected in the individual student's level but also in the entire campus's innovative environment. The least importance is given to the Performance in Innovation Competitions, indicating that the higher vocational innovative spirit education should depart from a competitive orientation which focuses mainly on 'competitions, papers and projects'.

4. Reliability and Validity Testing of the Evaluation Index System

The scientific soundness of the evaluation system is tested by student questionnaire data. Among the 2,500 student questionnaires distributed, 2,412 valid questionnaires were obtained after excluding the samples of abnormal response time and regular response, and the effective rate was 96.5%. Content validity was assessed by 8 independent experts who were not involved in the selection of indicators. The Item-level Content Validity Index of the 16 indicators was no less than 0.88, and the average Scale-level Content Validity Index was 0.96, above the general criterion of 0.90. The valid samples were randomly divided in half, and 1,206 of them were taken to complete the confirmatory factor analysis. All the goodness-of-fit indicators were within the acceptable range.

As shown in Table 5, all standardized factor loadings (0.69–0.88, $p < 0.001$), Composite Reliability values (0.86–0.90) and Average Variance Extracted values (0.58–0.65) satisfy their corresponding criteria, which confirms the convergent and discriminant validity. The overall scale has a Cronbach's α of 0.91, with subscale values varying from 0.85 to 0.90.

Table 5. CFA Results, Convergent Validity, and Reliability

Dimension / Indicator	Std. Loading	CR	AVE	α
B1 Reaction Level		0.87	0.58	0.87
C1 Content Satisfaction	0.72			
C2 Method Satisfaction	0.78			
C3 Instructor Satisfaction	0.81			
C4 Condition Support	0.69			
C5 Participation & Completion	0.80			
B2 Learning Level		0.86	0.61	0.85
C6 Theoretical Knowledge	0.75			
C7 Practical Skills	0.84			
C8 Innovation Awareness	0.74			
B3 Behavior Level		0.90	0.65	0.90
C9 Innovative Practice	0.73			
C10 Innovation & Entrepreneurship	0.82			
C11 Self-directed Innovation	0.88			
C12 Teamwork & Sharing	0.79			
B4 Results Level		0.89	0.61	0.88
C13 Campus Atmosphere	0.77			
C14 Student Innovation Output	0.85			
C15 Teacher Evaluation	0.70			
C16 Competition Performance	0.81			
Model Fit	$\chi^2/df = 2.18$, RMSEA = 0.048, CFI = 0.96, TLI = 0.95, GFI = 0.93			
Overall α	0.91			

Note: The significance level of all standardized loadings in the table reached $p < 0.001$. CR is the composite reliability, AVE is the average variance extracted, and α is Cronbach's alpha.$p < 0.001$

5. Conclusion

This study applies the Kirkpatrick Evaluation Model to evaluate the effectiveness of innovative spirit education in higher vocational colleges of Guangdong Province. A quantitative evaluation system with 4 levels of criteria and 16 levels of indicators is established, and the scientific weights and consistency are achieved by AHP method, resulting in three aspects of improvement: changing from the subjective qualitative assessment to a combination of quantitative and qualitative methods, progressing from single result assessment to an overall coverage of the process and results, and evolving from the scattered evaluation to a systematic and hierarchical evaluation.

The weight results show that the higher vocational innovative spirit education should follow three aspects: Firstly, based on the active participation of students, improve the provision and implementation methods of education. Secondly, take the cultivation of innovation awareness as the main point, enhance the formation of thinking and build up confidence. Thirdly, put emphasis on behavioral change, encourage the transformation of innovation awareness into practical actions. At the same time, the improvement of individual abilities and the establishment of an innovative campus environment should also be considered, constructing a lasting educational system that "cultivates awareness, trains skills, promotes behavior, and creates atmosphere."

6. Limitations and Future Research

Several limitations should be noted. First, the sample was confined to ten higher vocational colleges in Guangdong Province, purposively selected to cover Double High Plan, provincial demonstration, and ordinary institutions across cities, which limit generalizability. Second, the cross-sectional self-report design precludes causal inference and cannot track outcomes longitudinally. Kirkpatrick Level 4 results require longer-term data, so the current operationalization captures only perceived short-term effects. Third, self-reported behavior, such as self-directed innovation, is susceptible to social desirability bias. Future research should adopt longitudinal designs, objective measures, and multi-region samples to validate and extend these findings.

Funding

2024 Guangdong Provincial Demonstration Course of Ideological and Political Education in Curriculum for Higher Vocational Colleges, College Students' Career Development and Employment Guidance (No.KCSZ2024028); 2024 Guangdong Provincial Education Science Planning Project (Higher Education Special Project) “Research on Integrating the Spirit of Historical Initiative into University Ideological and Political Education” (No.2024GXJK095); Excellent Online Open Course & AI Demonstration Course Career Development and Employment Guidance for College Students (No.2026-JPKC-26);Planning Project of University Graduates Employment Association (2026) Research on the Path of Integrating Chinese Excellent Traditional Vocational Guidance Culture into College Students’ Career Development.

References

[1] Ministry of Education. Opinions on Accelerating the Reform of Innovation and Entrepreneurship Education in Higher Education Institutions in the New Era [Z]. 2018.
[2] Huang Z. X., Wang Z. Q. (2017) Disciplinary Characteristics and Development Paths of Innovation and Entrepreneurship Education in Universities. Educational Research, 2017(05), 46-54.
[3] Kirkpatrick, D. L. (1959). Techniques for evaluating training programs. Journal of the American Society of Training Directors, 13(11), 3–9.
[4] Hong Z. G, Li Y., Fan Z. H., Wang Y. (2002) Calculation of Higher-Order Average Random Consistency Index (RI) in Analytic Hierarchy Process. Computer Engineering and Applications, 12, 45-47.

[5] Meng S. M. (2021) The Value, Concepts, and Methods of Cultivating Innovative Competence among College Students in the New Era. China Higher Education, 17, 40-42.
[6] Zhou L. F., Wu Y. H. (2021) Multi-dimensional Approaches to Solidly Promoting the Evaluation of Higher Education Quality. Theoretical Research, 10, 50-56.
[7] Wu Y. G. (2021). A Study on the Integration Path of Cultivating Innovative Spirit and Professional Teaching for Higher Vocational Students. Vocational and Technical Education, 26, 36-40.
[8] Department of Education of Guangdong Province. The "14th Five-Year Plan" for the Development of Vocational Education in Guangdong Province [Z]. 2021.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261115

Multi-modal Digital Inheritance of Intangible Cultural Heritage: A Child-Friendly Micro-course Framework and Empirical Study—Taking the "Xiangnan Little Heritage Bearers" Digital Curriculum System as an Example

Yan Peng, Han Wang and Di Wang[1]
School of Education Science and Law, Xiangnan University, Chenzhou 423000, China

Abstract. In the context of rapid digital transformation, the traditional transmission of Intangible Cultural Heritage (ICH) is facing increasing challenges. Conventional approaches mainly depend on oral instruction and static cultural displays, which are often unable to meet the cognitive and interactive needs of contemporary children. To address this issue, this study proposes the "Xiangnan Little Heritage Bearers" framework, a child-oriented digital micro-course system integrating augmented reality (AR), digital reconstruction, and adaptive recommendation mechanisms. Using representative Xiangnan cultural resources such as Xiang-Kun Opera and Linwu Nuo Masks, the framework reconstructs traditional cultural content into immersive learning modules. A 24-week empirical experiment involving 320 primary school students demonstrates that the proposed system significantly improves cultural cognition, practical participation, and emotional engagement.

Keywords. Intangible Cultural Heritage; Augmented Reality; Digital Learning; Cultural Inheritance

1. Introduction

Intangible Cultural Heritage carries local historical memory, cultural identity, and collective social values [1]. However, with the acceleration of urbanization and the popularity of digital media, children's opportunities to participate in traditional cultural activities have gradually decreased. In many regions, the inheritance of local culture has become disconnected from children's everyday learning experiences.

At present, most heritage preservation projects focus on archival recording, museum exhibitions, or documentary production. Although these methods are important for cultural preservation, they often lack interactive participation. Compared with

[1] Corresponding Author: Di Wang, School of Education Science and Law, Xiangnan University, Chenzhou, China; Email: 1126330791@qq.com
This paper is a research outcome of the National College Students' Innovation and Entrepreneurship Training Program of China (Grant No. 202510545036).

traditional learning materials, contemporary children are generally more responsive to visualized, game-based, and immersive digital experiences [2][3].



Figure 1. Xiang-Kun Opera performance scene.

As shown in Figure 1, Xiang-Kun Opera contains rich visual symbols, costume aesthetics, and performative gestures. However, for young learners, long-duration opera performances may be difficult to understand through conventional classroom explanation alone. This limitation highlights the importance of transforming cultural resources into more accessible and interactive forms.

2. Research Methodology and System Development

2.1 The Experience–Empathy–Action (EEA) Model

To improve learning engagement and cultural understanding, this study proposes the Experience–Empathy–Action (EEA) pedagogical model. The model transforms static cultural materials into progressive immersive learning experiences. The theoretical foundation of this model is partially derived from constructivist learning theory and children's cognitive development theory proposed by Vygotsky and Piaget[4][5].

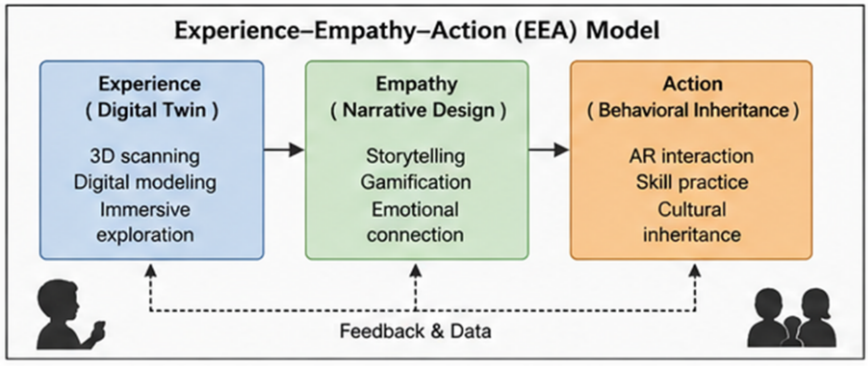


Figure 2. Workflow of the Experience–Empathy–Action (EEA) model.

As illustrated in Figure 2, the Experience stage focuses on digital perception through 3D scanning and virtual interaction. The Empathy stage reorganizes historical stories and ritual meanings into child-friendly narratives. Finally, the Action stage encourages children to participate in simulated cultural practices through AR interaction. This gradual process strengthens emotional resonance and behavioral participation.

2.2 Digital Reconstruction of Cultural Resources

Traditional ICH activities are usually long and structurally complex. Therefore, a micro-unit reconstruction strategy was adopted in this study. Large-scale cultural performances and rituals were divided into short interactive learning units lasting approximately 5–15 minutes.



Figure 3. 3D reconstruction of a Linwu Nuo mask.

Figure 3 presents the digital reconstruction process of a Linwu Nuo mask. The reconstruction workflow includes texture acquisition, geometric modeling, and lightweight rendering optimization. Physically Based Rendering (PBR) techniques were further applied to preserve visual authenticity while maintaining compatibility with mobile devices [6].

3. Technical Implementation and Algorithmic Optimization

3.1 AR Spatial Alignment and Interaction

To ensure stable interaction between virtual cultural objects and physical learning environments, an ORB-SLAM2-based spatial positioning method was employed[3]. Accurate spatial alignment is particularly important for young learners because excessive visual drift may negatively influence immersion and cognitive continuity.

$$E = \Sigma\Sigma\, w_{ij}\, \rho(||x_{ij} - \pi(K,R_j,t_j,X_i)||^2)$$



Figure 4. AI-generated illustration of children using AR-supported learning activities. No real child participants are depicted in this image.

As shown in Figure 4, children interact with reconstructed cultural artifacts through AR-supported tablet devices. Compared with passive video watching, this interaction mode enables learners to observe details, manipulate virtual objects, and complete cultural tasks in real time. Previous studies have shown that AR-based interaction can significantly improve learner engagement and immersion [2][7][8].

3.2 Adaptive Recommendation Algorithm

To reduce repetitive learning experiences and improve personalized instruction, this study developed a Hybrid Content-Based Recommendation (H-CBR) algorithm. The algorithm dynamically adjusts course difficulty according to learner profiles, behavioral feedback, and interaction frequency.

$$\mathrm{Sim}(U,C)= \Sigma(U_i \times C_i)/(\sqrt{\Sigma U_i^2} \times \sqrt{\Sigma C_i^2})$$

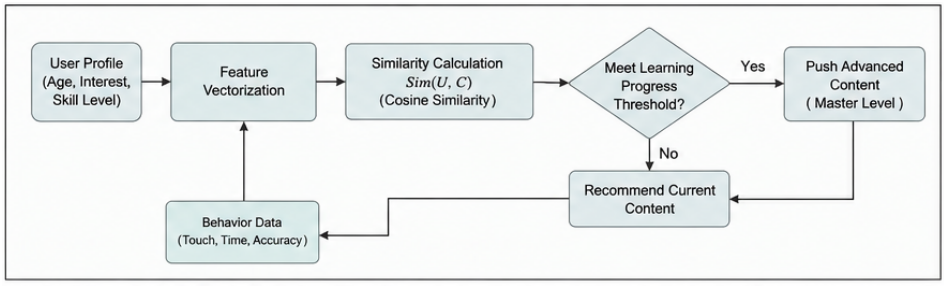


Figure 5. Workflow of the adaptive recommendation algorithm (H-CBR).

Figure 5 demonstrates the workflow of the adaptive recommendation process. The system continuously records user behavior data, including task completion time, interaction latency, and operational accuracy. When learners reach predefined performance thresholds, the system automatically unlocks more advanced cultural challenge modules.

4. Empirical Research and Analysis

4.1 Experimental Design

The empirical experiment lasted 24 weeks and involved 320 students from four primary schools in Xiangnan. A stratified random assignment procedure was used to ensure comparable distributions of age and gender between groups. Participants were randomly divided into a Control Group (CG) and an Experimental Group (EG). The Control Group received traditional video-based instruction, whereas the Experimental Group used the CFDI digital micro-course system.

Four assessment instruments were employed:

(1)Cultural Cognition Scale (CCS, 20 items);

(2)Practical Participation Scale (PPS);

(3)Heritage Engagement Questionnaire (HEQ, 5-point Likert scale);

(4)Classroom Observation Checklist (COC). Cronbach's alpha values were 0.84 and 0.88 for CCS and HEQ respectively.

The CCS and HEQ instruments were adapted from previous heritage education and cultural engagement studies. Data were analyzed using SPSS 27.0. Independent-sample t-tests and Cohen's d effect sizes were calculated. Statistical significance was set at $\alpha = 0.05$.

To ensure experimental consistency, both groups studied the same cultural content under similar teaching schedules. Evaluation indicators included cultural cognition, practical skill proficiency, behavioral participation, and emotional attitudes toward heritage preservation.

4.2 Quantitative Results and Behavioral Analysis

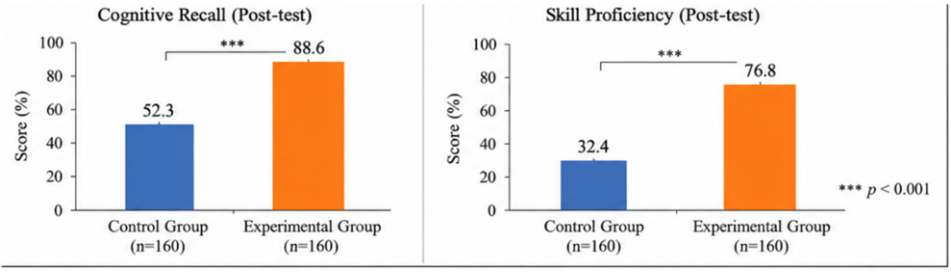


Figure 6. Comparison of learning outcomes between the control and experimental groups.

As shown in Figure 6, Students in the Experimental Group achieved significantly higher scores in both cognitive recall and practical skill proficiency. For cognitive recall, the Experimental Group achieved a post-test score of 88.6%, compared with 52.3% in the Control Group. For practical skill proficiency, the Experimental Group obtained a score of 76.8%, whereas the Control Group scored 32.4%. Independent-sample t-tests confirmed that both differences were statistically significant ($p < 0.001$).

Classroom observation records also showed that students using the CFDI system demonstrated stronger curiosity, more frequent collaborative discussion, and longer periods of sustained attention during learning activities. Teachers reported that AR-supported interaction helped students develop stronger emotional connections with

local culture. This finding is consistent with previous studies on immersive learning and flow theory [9].

4.3 Reliability and Educational Validity

To ensure research reliability, Cronbach's alpha coefficients were calculated for the cultural cognition and engagement scales. All coefficients exceeded 0.80, indicating acceptable internal consistency. In addition, triangulation methods combining quantitative assessment, classroom observation, and interview feedback were adopted to improve research validity.

5. Discussion

The findings suggest that immersive digital technologies can significantly improve the accessibility and attractiveness of ICH education for children. Digital reconstruction allows rare cultural resources to become more accessible, while adaptive recommendation mechanisms help maintain appropriate learning difficulty. The observed improvements may be explained by the combination of experiential learning and embodied interaction. Consistent with previous AR education studies [7,10–15], immersive and interactive learning environments promote active exploration, improve learner engagement, and facilitate deeper cultural understanding. Furthermore, the EEA model extends constructivist learning theory by integrating emotional resonance and cultural participation into digital heritage education.

At the same time, the study also highlights several practical concerns.

First, the sample was restricted to four primary schools in Xiangnan, which may limit generalizability.

Second, the intervention period lasted only 24 weeks and therefore cannot fully evaluate long-term cultural identity formation.

Third, the current system primarily focuses on visual and interactive learning experiences, while family and community participation were not included in the evaluation framework.

Future studies should incorporate longitudinal tracking and cross-regional comparative designs.

Fourth, although gamified interaction improves participation, excessive entertainment-oriented design may weaken cultural authenticity. Therefore, close collaboration with local inheritors remains essential throughout the content development process.

Fifth, hardware accessibility and teacher training remain important challenges for large-scale implementation. Future educational policies should therefore consider infrastructure support and interdisciplinary cooperation between schools, cultural institutions, and technology developers.

6. Ethics Statement

This study involved 320 primary school students.

According to institutional policy, formal ethical review was not required for this educational intervention study.

Written informed consent was obtained from participants' parents or legal guardians before data collection.
All participant data were anonymized and used exclusively for research purposes.

7. Conclusion

This study proposes a child-friendly digital inheritance framework integrating AR interaction, digital reconstruction, and adaptive recommendation technologies. By transforming static cultural resources into immersive learning experiences, the "Xiangnan Little Heritage Bearers" framework provides a practical approach for contemporary ICH education. Empirical evidence demonstrates that the framework significantly improves cultural cognition, practical participation, and emotional engagement among primary school students. Future research may further explore collaborative digital heritage creation and federated learning mechanisms to support long-term cultural sustainability.

References

[1] UNESCO. Convention for the Safeguarding of the Intangible Cultural Heritage. Paris: UNESCO, 2003.
[2] Azuma R. T. A Survey of Augmented Reality. Presence: Teleoperators and Virtual Environments, 1997, 6(4): 355–385. DOI:10.1162/pres.1997.6.4.355.
[3] Mur-Artal R., Tardós J. D. ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras. IEEE Transactions on Robotics, 2017, 33(5): 1255–1262. DOI:10.1109/TRO.2017.2705103.
[4] Vygotsky L. S. Mind in Society: The Development of Higher Psychological Processes. Cambridge, MA: Harvard University Press, 1978.
[5] Piaget J. The Psychology of the Child. New York: Basic Books, 1969.
[6] Wang Y., Gu T., Wu M. Digital Preservation and Interactive Dissemination of Intangible Cultural Heritage in China. Heritage Science, 2021, 9(1): 88–101. DOI:10.1186/s40494-021-00528-3.
[7] Bacca J., Baldiris S., Fabregat R., Graf S., Kinshuk. Augmented Reality Trends in Education: A Systematic Review of Research and Applications. Educational Technology & Society, 2014, 17(4): 133–149.
[8] Kirner C., Kirner T. G. Virtual Reality and Augmented Reality Applied to Simulation Visualization. Simulation & Gaming, 2011, 42(4): 523–544. DOI:10.1177/1046878110372642.
[9] Csikszentmihalyi M. Flow: The Psychology of Optimal Experience. New York: Harper & Row, 1990.
[10] Ibáñez M. B., Delgado-Kloos C. Augmented Reality for STEM Learning: A Systematic Review. Computers & Education, 2018, 123: 109–123. DOI:10.1016/j.compedu.2018.05.002.
[11] Akçayır M., Akçayır G. Advantages and Challenges Associated with Augmented Reality for Education: A Systematic Review of the Literature. Educational Research Review, 2017, 20: 1–11. DOI:10.1016/j.edurev.2016.11.002.
[12] Wu H. K., Lee S. W. Y., Chang H. Y., Liang J. C. Current Status, Opportunities and Challenges of Augmented Reality in Education. Computers & Education, 2013, 62: 41–49. DOI:10.1016/j.compedu.2012.10.024.
[13] Dunleavy M., Dede C., Mitchell R. Affordances and Limitations of Immersive Participatory Augmented Reality Simulations for Teaching and Learning. Journal of Science Education and Technology, 2009, 18(1): 7–22. DOI:10.1007/s10956-008-9119-1.
[14] Billinghurst M., Clark A., Lee G. A Survey of Augmented Reality. Foundations and Trends in Human–Computer Interaction, 2015, 8(2–3): 73–272. DOI:10.1561/1100000049.
[15] Champion E. Critical Gaming: Interactive History and Virtual Heritage. Farnham: Ashgate Publishing, 2015.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)

© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261116

Research and Practice on the Construction of a Lightweight AI-Enabled Welding Experimental Teaching and Safety Intelligent Management System

Haifeng Yang[a,b], Gang Chen[a,b], Meirong Wang[a,b1] and Xiongyao Wang[a,b]
[a] *Key Laboratory of Special Welding Technology Shandong Province, Harbin Institute of Technology, Weihai, Weihai 264209, China*
[b] *State Key Laboratory of Precision Welding & Joining of Materials and Structures, Harbin Institute of Technology, Harbin, 150001, China*

Abstract. Aiming at the four prominent problems existing in current multi-process welding laboratories, including difficult supervision of teaching processes, high pressure on safety risk prevention and control, fragmented integration of teaching and management, and difficult implementation of intelligent transformation, this paper takes an eight-category multi-process, multi-equipment, and multi-scene welding experimental teaching system as the research object. Adopting the methods of literature research, field investigation, pilot verification and closed-loop optimization, this study constructs an integrated intelligent management system with in-depth integration of teaching and safety, covering the whole experimental process of pre-class, in-class and after-class stages supported by lightweight artificial intelligence (AI) technology. Lightweight AI capabilities are deeply embedded in four key links: standardization of teaching procedures, standardization of operational behaviors, evaluability of experimental quality, and early warning of safety risks, forming a low-cost, easy-to-implement, replicable and scalable intelligent management paradigm for engineering experiments. Practical results show that the proposed system significantly improves the standardization, proceduralization and objectivity of welding experimental teaching, strengthens the capabilities of pre-event early warning, rapid disposal and closed-loop rectification of safety risks, and effectively reduces human supervision omissions. It provides a referable practical path for the digital transformation of university experimental teaching and the modernization of laboratory governance capacity.

Keywords. Lightweight AI; welding experiment; experimental teaching; safety management; intelligent laboratory; emerging engineering education

1 Introduction

As a core module of the practical teaching system for engineering majors, welding experiments serve as a key carrier for cultivating students' engineering practical

[1] Corresponding Author: Meirong Wang, Email: wmr21c@163.com

abilities, process standard awareness and safety production literacy [1-3]. This study defines two quantifiable expected learning outcomes for students supported by the proposed system: Process cognition ability: ①Students can select appropriate welding parameters according to different working conditions, identify basic weld defects, and understand quantitative evaluation indicators for practical training; ②Comprehensive engineering competence: Students can complete standardized experimental reports based on AI training data, and form engineering thinking covering process review, process optimization and risk prediction.

The traditional experimental management mode relying on manual supervision, empirical evaluation and post-event rectification can no longer adapt to the needs of modern engineering practical teaching [4,5]. Welding experiments cover a variety of high-risk processes such as manual welding, laser welding, additive welding and non-destructive testing, featuring complex equipment types, diverse operational scenarios and concealed safety hazards, which makes them one of the practical teaching categories with the highest safety risks and the most difficult teaching supervision in university laboratories [6-9].

At present, most multi-process welding laboratories in domestic universities are confronted with numerous practical problems that severely restrict the improvement of practical teaching quality and the upgrading of laboratory safety governance. First, the coverage of teaching process supervision is insufficient. Thus, the standardization of student operation, completeness of experimental procedures and authenticity of practical training data are difficult to trace and quantitatively evaluate [10]. Second, the safety risk prevention and control system is backward. Welding practical training involves multiple risks such as high temperature, strong light, arc radiation, flammability and explosibility, and mechanical damage. Manual inspection is limited by hysteresis, subjectivity and locality, making it impossible to capture and early warn against irregular operations and potential safety hazards in real time, resulting in prominent pressure on safety accident prevention and control [11, 12]. Third, the cost of intelligent transformation is excessively high. Most existing intelligent laboratory management solutions rely on customized software development and overall hardware renovation, which are characterized by high investment cost, long construction cycle and high technical threshold. Ordinary experimental teachers cannot independently implement and promote such solutions, which makes it difficult to adapt to the lightweight upgrading needs of welding laboratories in ordinary universities [14-16].

This study abandons the traditional high-cost and high-threshold intelligent transformation mode, and deeply integrates AI technology into the whole process of pre-class preview, in-class management and after-class review of welding experiments based on existing monitoring equipment, teaching resources and mature lightweight AI tools. An integrated intelligent management system with dual empowerment of teaching quality improvement and safety prevention and control is constructed to solve the core problems of difficult supervision, weak safety management and difficult intelligent implementation in multi-process welding experimental teaching. This research constructs a low-cost, replicable and easy-to-promote intelligent teaching and safety management model for engineering experiments, providing practical references and paradigm support for the digital transformation of university experimental teaching and the modernization of laboratory safety governance [17-18].

2 Research Status at Home and Abroad

Developed countries in Europe and America have long been committed to the intelligentization of engineering practical training and the intellectualization of laboratory safety management. Relying on cutting-edge technologies such as artificial intelligence, computer vision, edge computing and VR/AR, they have formed development characteristics of lightweight technology, scenario-based application and closed-loop management, and accumulated rich practical experience in AI-enabled intelligent teaching and safety management of welding practical training [19-21].

In terms of intelligent practical teaching research, foreign scholars focus on the intelligent training and process evaluation of welding operation. They integrate deep learning and computer vision technology with welding process practical training scenarios to realize the segmentation of welding operation actions, discrimination of operation standardization, evaluation of process quality and real-time adaptive guidance, which effectively solves the problems of untimely teacher guidance and lack of personalized teaching in traditional welding practical training [22-23]. Meanwhile, some studies build immersive welding practical training platforms combined with VR virtual simulation technology, avoiding potential safety risks in physical practical training, greatly improving the efficiency of practical teaching and students' operational proficiency, and providing a new approach for high-risk welding practical teaching [24].

Closely focusing on the policy requirements of emerging engineering construction and digital education transformation, domestic academic circles have continuously carried out research on intelligent laboratory construction, AI-assisted experimental teaching and intelligent laboratory safety management, achieving phased research results and laying a theoretical foundation for the intelligent reform of engineering experiments [3].

In the field of special research on welding experiments, existing studies are mostly limited to single dimensions such as intelligent identification of welding defects, optimization of welding process parameters and virtual welding simulation training, which only solve the problems of welding quality control and early-stage practical training [18]. Nevertheless, significant research gaps remain in current domestic studies. First, the research scenarios are single, most studies focus solely on teaching quality improvement or safety management, lacking an integrated system with in-depth integration of teaching and safety. Second, the practicality is insufficient. Most studies rely on customized system development and hardware renovation, which are high-cost and difficult to promote, and cannot be independently implemented by front-line experimental teachers. Third, the adaptability is poor. Existing studies fail to fit the complex welding practical training scenarios of multi-process, high-risk and multi-station coexistence in universities, and cannot fully cover the teaching and safety management needs of eight mainstream welding experiments.

In summary, there is a lack of lightweight, low-cost and implementable AI-enabled integrated teaching and management solutions adapted to multi-process welding experimental scenarios in domestic universities. Based on the shortcomings of existing research, this study constructs a full-process closed-loop intelligent management system for welding experimental teaching and safety by virtue of mature lightweight AI tools, which has outstanding theoretical innovation and practical applicability.

3 Research Ideas and Methods

This study adheres to the core research logic of "problem orientation-practice driven-lightweight AI-result orientation". Taking the real teaching and safety management scenarios of multi-process welding experiments in universities as the carrier, the research is carried out following the technical route of "current situation sorting-framework construction-module implementation-data acquisition-system integration-optimization and upgrading-result condensation". Taking the whole process of multi-category welding experimental teaching as the core research object, this study disassembles the key links of teaching implementation, safety management and quality evaluation. Lightweight AI tools are adopted as the core technical means to realize intelligent recognition of operational behaviors, intelligent analysis of practical training videos, procedural evaluation of teaching quality and auxiliary decision-making of safety management. Standardized system construction is taken as the guarantee to solidify the AI application mode into normalized teaching and management specifications, so as to achieve the collaborative improvement of welding experimental teaching quality and laboratory safety management level with closed-loop quality improvement as the ultimate goal.

Four core research methods are adopted in this study. First, the literature research method. This study systematically sorts out high-level literature in the fields of emerging engineering experimental teaching, AI intelligent laboratories, intelligent welding practical training and laboratory safety management in recent three years, summarizes the achievements and deficiencies of existing research, clarifies the research innovation points, and avoids repetitive research. Second, the investigation and analysis method. Through on-site laboratory investigation, interviews with front-line experimental teachers and student questionnaire surveys, the core pain points and difficult problems in multi-process welding experimental teaching and safety management are accurately identified, and special investigation reports are formed to provide practical basis for system construction. Third, the system construction method. Combined with literature research and investigation results, the core modules, technical architecture and operation process of the intelligent management system are determined, the system architecture diagram is drawn, and expert demonstration and optimization are conducted in the fields of educational technology, welding technology and laboratory management to form a scientific and feasible implementation plan. Fourth, the pilot verification method. A three-dimensional comprehensive evaluation framework including quantitative index comparison, qualitative questionnaire & interview, and long-term closed-loop tracking is constructed to fully verify the effectiveness of the proposed system:

① Quantitative evaluation: Set control groups adopting traditional manual management and experimental groups adopting the proposed AI intelligent system. Collected indicators include frequency of non-standard student operations, number of safety violation records, average training scoring, teachers' management working hours, and quantity of undetected hazards. Statistical comparison is adopted to quantify practical effects; ② Qualitative evaluation: Distribute student satisfaction questionnaires and teacher interview outlines to collect subjective feedback on the usability of AI tools, teaching improvement and safety management optimization; ③ Long-term closed-loop evaluation: Track hazard rectification closed-loop rate, quality of student experimental reports and declining trend of repeated violations for two consecutive semesters to verify

long-term stable performance of the system. ④ Undergraduate classes are selected for teaching pilot tests, the implementation effects of the AI intelligent management mode and the traditional manual management mode are compared, teacher and student feedback data are collected, system functions and operation mechanisms are continuously optimized, and a implementable and promotable implementation paradigm is formed.

Mature commercial lightweight AI video analysis tools based on lightweight object detection model and time-series action recognition algorithm are applied to record standard operation templates. The system compares real-time training footage to identify students' irregular operations and generates problem lists to support after-class review and rectification. An open-source lightweight industry large language model fine-tuned on welding training datasets is adopted to build modules for experimental report evaluation and intelligent welding process guidance, and the model recognition accuracy and process adaptability are continuously iterated and optimized.

Secondly, all practical training stations in the laboratory are comprehensively investigated, divided into high-risk and general-risk levels, and all safety risks considered and integrated into the AI early warning system are systematically sorted out as follows:

① High-temperature burn risk: Scalding caused by contact with welding molten pool, high-temperature workpieces and hot welding torch components; ② Arc radiation risk: Ocular electro-optic inflammation and skin burns from unprotected direct viewing of welding arcs; ③ Missing personal protective equipment risk: Violations including absence of welding helmet, protective gloves and flame-retardant overalls; ④ Flammable and explosive risk: Ignition of surrounding oil stains, paper drawings and combustible supplies by welding sparks;

The above 4 types of risks are classified into high-risk items (arc radiation burns, flammable and explosive hazards, improper live equipment operation) and general-risk items (missing protection, welding fumes, mechanical collision). Core hidden hazards and AI visual identification priorities of each station are summarized to form a hierarchical risk management list. Based on existing monitoring equipment, AI visual recognition tools are connected, and irregular recognition rules, early warning thresholds and disposal methods are configured to establish a safety closed-loop management mechanism of "AI recognition-real-time early warning-on-site disposal-archive record-review and rectification".

4 Innovative Research Characteristics

4.1 System Innovation: Constructing a New Integrated Intelligent Governance Model of Teaching and Management

Breaking the fragmented governance mode of separated welding experimental teaching and safety management, this study constructs a dual-integrated intelligent management system of "teaching quality improvement + safety prevention and control" empowered by lightweight AI. The system is fully adapted to multi-process composite practical training scenarios including manual arc welding, laser welding, additive welding and non-destructive testing, covering the whole process, all scenarios and all personnel of experiments. It realizes the standardization of teaching evaluation, intellectualization of safety management and closed-loopization of process management, filling the research

gap of integrated and intelligent teaching and management for multi-process welding experiments.

4.2 Path Innovation: Proposing a New Low-Cost Lightweight Implementation Path

Different from the traditional high-cost and high-threshold customized development and hardware renovation mode of intelligent laboratories, this study adopts mature commercial lightweight AI tools without special software development and hardware equipment renovation. It can complete intelligent upgrading with low cost, zero technical threshold and short cycle. The scheme adapts to the laboratory construction budget of ordinary universities and the operational ability of front-line experimental teachers, supporting independent deployment and rapid implementation, and thoroughly solving the problems of difficult implementation and promotion of traditional intelligent schemes.

4.3 Mechanism Innovation: Establishing a New Full-Process Closed-Loop Improvement Mechanism

This study reconstructs the application logic of AI technology, upgrading AI from a single safety monitoring tool to a core supporting tool covering the whole chain of experimental teaching, safety management, quality evaluation and review optimization. Through real-time data collection, intelligent analysis, accurate early warning and quantitative evaluation of AI, a closed-loop operation mechanism of "practical training implementation-intelligent monitoring-problem identification-rectification and optimization-review and improvement" is formed, realizing the continuous iterative optimization of teaching quality and safety management.

5 Research Achievements and Practical Effects

(1) Compilation of complete standardized documents for teaching and safety management

Combined with the process characteristics and safety specifications of eight types of welding experiments, a complete set of institutional achievements has been compiled, including the *Standard Operation Manual for Multi-Process Welding Experiments*, *Quantitative Evaluation Standards for Experimental Processes*, *Hierarchical List of Laboratory Safety Risks*, *Specifications for AI Intelligent Early Warning Disposal* and *System Implementation Rules*. The AI intelligent management mode is solidified into standardized, assessable and implementable normalized management processes, realizing the standardized and normalized operation of welding experimental teaching and safety management.

(2) Deployment of lightweight AI intelligent functional modules to realize full-process intelligent management of practical training

Based on mature lightweight AI tools, five core functional modules are deployed, including intelligent operational behavior recognition, practical training video intelligent analysis, real-time operational quality scoring, intelligent safety violation early warning, and auxiliary experimental report evaluation. The system realizes

automatic recognition of students' irregular operations, full-process traceability of practical training, quantitative and objective teaching evaluation, and real-time early warning of potential safety hazards, thoroughly solving the problems of insufficient manual supervision coverage, subjective evaluation and delayed hazard detection in traditional modes.

(3) Significant pilot application effects with greatly improved teaching and safety levels

Pilot applications are carried out in undergraduate classes majoring in mechanical and material engineering. Practical data show that after the application of the system, the standardization of students' welding experimental operations is significantly improved, the frequency of irregular operations is greatly reduced, and the standardization, proceduralization and objectivity of experimental teaching are comprehensively enhanced. The safety violation rate and hazard omission rate of laboratories are significantly decreased, and the ability of pre-event prevention and control of safety risks is remarkably improved. Teachers' workload in manual supervision, on-site evaluation and hazard inspection is greatly reduced, effectively releasing teachers' teaching energy. The system achieves multiple goals of improving welding experimental teaching quality, enhancing safety efficiency and reducing management burden.

6 Conclusion

This paper constructs a full-coverage, full-process and closed-loop integrated system for welding experimental teaching and safety management based on lightweight AI technology. Breaking the dual limitations of traditional manual management and traditional heavy-asset intelligent transformation, the system realizes the dual empowerment of standardized improvement of welding experimental teaching quality and precise management and control of laboratory safety with the advantages of low cost, lightweight, easy implementation and replicability, effectively making up for the shortcomings of intelligent management of multi-process welding experiments.

References

[1] Hu Q T. Research on the theoretical framework and practical path of artificial intelligence empowering university experimental teaching[J]. China Educational Technology, 2025(3):56-64.
[2] Yang X. A comparative study on the application of AI experimental teaching at home and abroad[J]. Chinese Journal of Higher Education, 2024(8):78-84.
[3]. Research on the innovation of university experimental teaching evaluation system empowered by artificial intelligence[J]. Research in Higher Education of Engineering, 2025(S1):1-5.
[4] Wang F, Zhang L C, You T T. Analysis and prospect of university teaching laboratory safety management in the digital intelligence era[J]. Laboratory Science, 2024,27(4):1-6.
[5]. Research and practice on intelligent teaching management of university laboratories under digital education[J]. Modern Educational Technology, 2025(1):89-95.
[6]. Exploration on future experimental teaching mode and intelligent autonomous laboratory construction[J]. Research and Exploration in Laboratory, 2026,45(3):1-7.
[7] Liu C F, Wang R. Application, challenges and countermeasures of artificial intelligence in laboratory safety management[J]. Experimental Technology and Management, 2025,42(4):239-244.
[8] Yang G X, Zhang X R, Zhu M. Research on closed-loop management system of laboratory safety hazards based on AI agent[J]. University Chemistry, 2025,40(12):1-8.
[9] Zhang J, Luo M R. Multi-dimensional exploration of AI agent empowering university laboratory management capability improvement[J]. Experimental Technology and Management, 2025,42(5):1-7.

[10] Chen W, He J. Optimization strategy of university laboratory safety management integrated with AI agent[J]. China Broadband, 2024(9):148-150.
[11]. Research on intelligent recognition of laboratory unsafe behaviors based on AI vision[J]. Research and Exploration in Laboratory, 2025,44(2):1-6.
[12] Li G. Research on safety risks and intelligent management of multi-process welding laboratories in universities[J]. Welding Technology, 2025,54(2):89-93.
[13] Zhao J. Application of lightweight AI in safety management and control of engineering training laboratories[J]. Experimental Technology and Management, 2025,42(6):198-202.
[14] Zhou M. Practice of digital transformation of welding experimental teaching under the background of emerging engineering[J]. China Metallurgical Education, 2025(2):67-70.
[15] Wu H. Construction of procedural evaluation system for welding practical training under the background of intelligent laboratory[J]. Exploration of Higher Vocational Education, 2025,24(1):56-61.
[16]. Research on AI-assisted teaching and quality control of multi-process welding practical training[J]. Chinese Vocational and Technical Education, 2025(10):1-6.
[17] Lassiter T, Collier C, Fleischmann K R. Welding Instructors' Perspectives on Using AI Technology in Welding Training[J]. Proceedings of ASIS&T, 2023,60(1):1-10.
[18] Querfurth B V, Belhout S E, Knaak C, et al. AI-driven autonomous adaptive feedback welding machine[J]. Welding in the World, 2025,69:1419-1426.
[19]. Deep learning-based action recognition for joining and welding processes of dissimilar materials[J]. Frontiers in Materials, 2025,12:1-12.
[20]. Integrating Virtual Reality into Welding Training: An Industry 5.0 Approach[J]. Electronics, 2025,14(10):1-18.
[21]. AI-powered laboratory safety monitoring and early warning system[J]. IEEE Access, 2024,12:1-10.
[22]. Lightweight deep learning for real-time action recognition in educational laboratories[J]. Sensors, 2025,25(8):1-16.
[23]. Intelligent evaluation system for experimental teaching based on multi-modal data fusion[J]. Computers & Education, 2024,15:1-14.
[24]. Smart laboratory management: A systematic review and future directions[J]. Journal of Laboratory Automation, 2025,30(2):1-12.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)

© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261117

Research on the Construction and Practice of AI Literacy Course

Xiaoli Liu[a], Wenqing Li[b], Jingjing Li[c], Zhen Zhang [a], Wei Li[b,*] and Anjia Yang [d,*]
[a] *College of Information Science and Technology, Jinan University, Guangzhou, China*
[b] *Undergraduate School, Jinan University, Guangzhou, China*
[c] *Guangxi Key Laboratory of Trusted Software, Guilin University of Electronic Technology, Guilin, China*
[d] *College of Cyber Security, Jinan University, Guangzhou, China*
ORCID: Xiaoli Liu https://orcid.org/0000-0001-6488-9732

Abstract. Amid rapid generative AI growth, cultivating AI literacy among non-computer science (non-CS) undergraduates is a key higher education reform priority. Drawing on national policies and domestic course experience, this study constructs a five-dimensional progressive AI literacy framework: fundamental knowledge, risk identification, safety awareness, ethical analysis, governance participation. Pedagogically, the course uses an online smart platform for multimodal delivery, case-based teaching, and integrated process-team assessment. In a teaching practice of AI Safety, Ethics, and Governance with 66 non-CS undergraduates (44 valid surveys), results show a moderate positive correlation between classroom interaction and team project quality; students highly value innovative assessments like team projects and self-designed questions, and most gain achievement from work selected as teaching cases. This study offers an empirically grounded, data-driven course design and evaluation framework, filling the general education gap for non-CS students.

Keywords. AI literacy, General education, Course design, Case-based teaching

1. Introduction

1.1 Technological Development and the Practical Demand for AI Literacy

Over the past decade, artificial intelligence has undergone a transformative evolution from deep learning to large language models, and generative AI is reshaping the ways information is produced and decisions are made. The ManpowerGroup 2026 global talent shortage survey (covering 41 countries and 39,000 employers) shows that AI model and application development (20%) and AI literacy (19%) rank for the first time as the top two most difficult skills to fill worldwide; in the UK specifically, 19% of employers cited AI as the hardest skill to recruit [1], [2]. In February 2026, the U.S.

* Corresponding Authors, Wei Li, Email: jnuliwei@126.com and Anjia Yang, Email: anjiayang@gmail.com
Xiaoli Liu and Wenqing Li contributed equally to this work.
This paper is supported by Guangxi Key Laboratory of Trusted Software (KX202306), National Natural Science Foundation of China (Project No.72574084) and Guangdong Province Education Science Planning Project (2024GXJK218).

Department of Labor released its AI Literacy Framework, defining AI literacy as "a set of capabilities that enable individuals to use and evaluate AI technologies responsibly and effectively" and explicitly identifying five core content areas: understanding AI principles, exploring AI applications, instructing AI effectively, evaluating AI outputs, and using AI responsibly. A subsequent training notice from the Department further states that AI literacy is becoming a foundational workforce skill, and that workers in nearly all industries will need basic AI knowledge [3].

The academic community has paid close attention to this issue. Long and Magerko (2020), in their CHI paper, proposed that AI literacy encompasses 17 competency elements, emphasizing that individuals should be able to critically evaluate AI technologies and communicate and collaborate effectively with AI [4]. Ng et al. (2021), in "Conceptualizing AI literacy: An exploratory review" published in Computers and Education: Artificial Intelligence, proposed a four-dimensional model consisting of "knowing and understanding — using and applying — evaluating and creating — ethical issues" [5]. However, empirical research on AI literacy courses designed for general education contexts targeting non-computer science university students remains limited.

1.2 Teaching Objectives of AI Literacy General Education Courses

In 2024, the Ministry of Education launched a demonstration initiative for the application of large language models in education systems, explicitly calling for the "development of a systematic AI general education curriculum system" and promoting project-based, scenario-based, and inquiry-based teaching approaches [6]. Against this background, the teaching objectives of this course are set to enable students to understand basic principles of AI, identify risks in AI applications, establish safety awareness, conduct ethical analysis, and participate in governance discussions, thereby cultivating comprehensive competencies for adapting to an intelligent society. These objectives respond to national policy directions and align with industry demands for integrated AI literacy competencies.

1.3 Structure of This Paper

This paper consists of five parts. Section 2 elaborates on the course content framework. Section 3 introduces the pedagogical implementation strategies. Section 4 presents teaching practice and data analysis using the course "AI Safety, Ethics, and Governance" as an example. Section 5 summarizes the research conclusions and discusses future directions.

2. Content Framework Development Approach

2.1 Investigative Analysis

2.1.1 Learning Needs Survey

To accurately capture students' learning motivations, a pre-course survey was conducted among the cohort of 2025 students via an online survey platform (Questionnaire Star), yielding 779 valid responses.

The main motivations for students to enroll in the AI general education course are shown in Fig 1: enhancing employment competitiveness (30.94%), empowering their own majors (14.51%), and understanding how AI changes daily life (13.74%). This indicates that students are more concerned with the practical benefits of AI literacy for their careers and majors, providing a basis for course design.

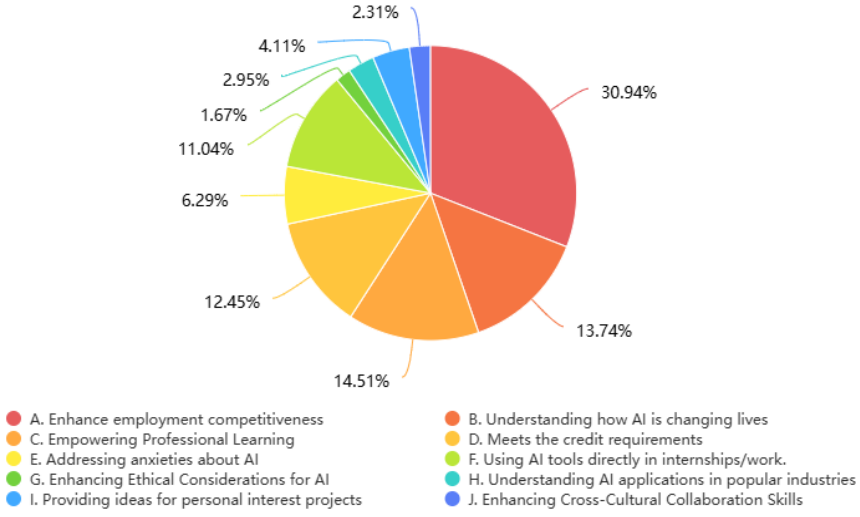


Fig 1. Students' enrollment motivations for the AI general education course

We use Rain Classroom, an AI-powered online teaching platform, to intelligently manage classroom instruction and online learning through features such as automated attendance, real-time quizzes, learning behavior analytics, and personalized resource recommendations[7].

Concerning the survey on learning resource forms, as shown in Fig 2, the majority of students (68.55%) preferred Rain Classroom or learning platform courseware, followed by a platform-exclusive resource library (45.96%), WeChat official account push notifications (38.51%), an electronic course manual (37.48%), and student-generated learning materials (35.82%). Only 8.34% indicated no need for additional resources.

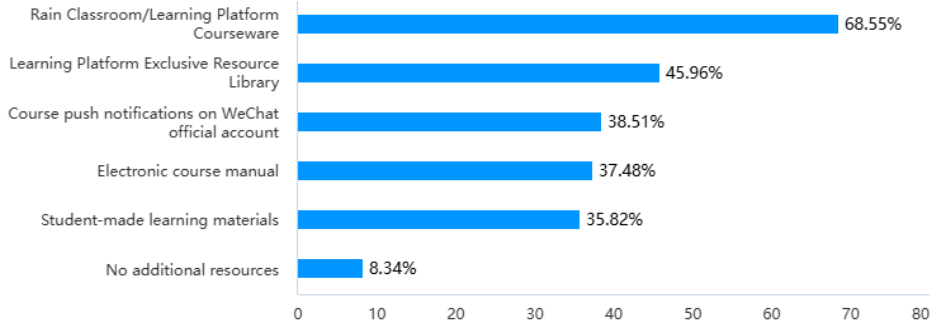


Fig 2. Students' preferences for learning resource formats

2.1.2 Current State of AI General Education Course Development in Domestic Universities

Prominent domestic universities have accelerated the development of AI general education courses. Tsinghua University's Department of Automation offers the general education elective course "Approaching Artificial Intelligence", which covers topics such as search, computer vision, machine learning, and trustworthy AI. It does not require programming and emphasizes the formation of systematic understanding through case studies[8],[9]. Fudan University has launched the AI-BEST course sequence, in which "Approaching Artificial Intelligence" is designed for students with no prior background, covering topics from basic concepts to large models and embodied AI, and incorporates hands-on practice [10]. As of the 2024–2025 academic year, Fudan University has offered 121 AI mega-courses covering all first-level disciplines [11]. The white paper issued by the Chinese Association for Artificial Intelligence proposes that AI general education should encompass four levels: "understanding AI", "using AI", "innovating with AI", and "governing AI well" [12]. Building on these institutional practices, this study further focuses on the systematic design of courses that emphasize the dimensions of safety, ethics, and governance.

2.2 Course Content Module Design

The white paper of the Chinese Association for Artificial Intelligence explicitly states that the content of an AI general education curriculum should include four organic components: "understanding AI", "using AI", "innovating with AI", and "governing AI well". This system clearly reflects an iterative and progressive logic from knowledge acquisition to competency development and from competency cultivation to literacy formation. Drawing on this framework, as well as the provisions of the EU DigComp 2.2 digital competence framework regarding citizen interaction with AI systems and China's "Ethical Norms for the New Generation Artificial Intelligence", this study organizes course content into five progressive modules. The modules follow a progressive pathway of "foundation → risk perception → safety response → ethical judgment → governance participation".

3. Pedagogical Implementation Strategies

3.1 Content Presentation

This course uses an online smart teaching platform for multimodal presentation of learning content. The platform supports a phased learning process across pre-class, in-class, and post-class stages, each with distinct pedagogical functions and illustrative examples.

Before class, the platform delivers micro-videos, foundational readings and pre-class quizzes to activate students' prior knowledge and establish a shared knowledge base. During class, dynamic knowledge graphs, live comments and real-time polling are adopted to stimulate classroom participation, and instructors can adjust teaching focus based on instant feedback to turn lectures into responsive discussions. After class, personalized scenario-based interactive exercises with immediate feedback are assigned to reinforce key concepts; the platform automatically records process data such as video completion rates, interaction frequency and quiz scores for teaching evaluation, shifting the teaching model from "experience-driven" to "data-driven" smart learning.

3.2 Case-Based Teaching

AI literacy courses are characterized by interdisciplinary content, context dependence, and value judgment, making traditional one-way lecturing insufficient for cultivating contextualized judgment abilities. Case-based teaching (CBT) offers several distinct advantages that align with the pedagogical needs of AI literacy education.

First, CBT connects abstract ethical principles with concrete AI scenarios, using real cases like the COMPAS recidivism algorithm controversy to make technical and moral concepts tangible and boost learning retention. Second, it cultivates critical thinking and perspective-taking through multi-stakeholder case debates, mirroring the collaborative governance requirements of real-world AI systems. Third, it reduces cognitive load for non-specialist learners via just-in-time concept exposure, and the course adopts a four-step weekly case procedure (pre-reading, group discussion, whole-class debate, instructor summary) to lower the learning threshold.

3.3 Assessment Design

Competence-oriented assessment requires observing behavioral performance in authentic contexts [13]. However, the AI era introduces fundamental changes in assessment philosophy and practice.

First, assessment shifts from product evaluation to process evaluation, tracking students' documented AI interactions to foster responsible AI use. Second, it incorporates authentic collaborative tasks, requiring teams to deliver analytical reports and presentations on real AI ethics or governance issues, with rubrics focusing on critical thinking, stakeholder analysis and feasible mitigation strategies. Third, it promotes transparency and student agency via co-designed assessment criteria, peer evaluation, and a well-received optional self-authored quiz activity.

4. Teaching Practice Case Study

4.1 Course Implementation and Process Data

This section presents the actual offering of the course "AI Safety, Ethics, and Governance". The course was designed according to the framework described above. Sixty-six non-computer science undergraduate students elected the course, which lasted 16 class hours and covered the stages of fundamental AI knowledge, risk identification, security technologies, ethical analysis, and governance frameworks. The content schedule is shown in table 1. These surveyed topics adopt a societal-impact-oriented approach to guide students to grasp specialized concepts from a general literacy standpoint instead of mathematical analysis.

Table 1 Content of "AI Safety, Ethics, and Governance"

No.	Title	Content
Lecture 1	Overview of AI risks	Core AI risk categories & real-world impacts
Lecture 2	AI fundamentals	Basic concepts & working principles of modern AI
Lecture 3	Agent Safety	Safety challenges & mitigation for autonomous AI agents
Lecture 4	AI Safety engineering	Technical methodologies & safe AI engineering practices
Lecture 5	AI Ethics	Ethical frameworks & moral dilemmas in AI deployment
Lecture 6	AI governance	Regulatory policies & AI oversight mechanisms

In this course, all 66 online learners participated (100% active rate). Students completed an average of 13.6 learning activities, with an overall completion rate of 80%, indicating good learning progress. Among them, 44 students (67%) used AI to assist their learning, with an average of 6.3 questions posed per student. AI tools have become a significant aid for student learning, providing effective support for interaction and personalized learning, and demonstrating the high degree of AI integration and student engagement in the course. Online learning data see Fig 3.



Fig 3. Learning Data from Rain Classroom Online AI Courses

4.2 Survey Results

A post-semester questionnaire for AI Safety, Ethics, and Governance yielded 44 valid responses, covering assessment methods, collaborative teamwork and outcome-based learning achievement, each illustrated in respective figures. As shown in Figure 4, nearly 60% of students (59.09%, combining "Absolutely necessary" and "Quite necessary") held that AI general education courses should break single-semester limits and recognize cross-semester continuous learning outcomes, while over 90% (90.91%) deemed the current open-book assessment reasonable.

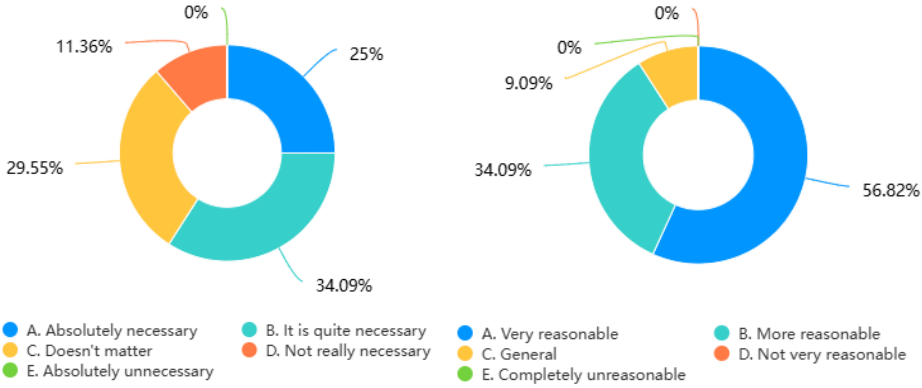


Fig 4. Survey on “Cross-Semester Assessment Necessity” and “Open-Book Final Exam Reasonableness”

Regarding learning styles and the survey results in Fig 5, nearly 70% (68.18%) of respondents endorsed the assessment model combining "individual learning performance + team collaboration outcomes". While 84.09% believed that the assessment of course team projects should pay attention to the social value and real-world significance of the projects, reflecting students' emphasis on collaborative learning and the practical impact of projects.

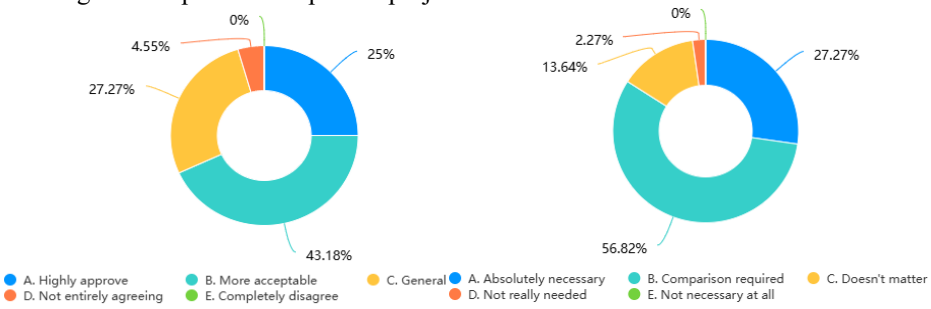


Fig 5. Survey on "Individual-Team Combined Assessment Approval" and "Team Project Social Value Focus Necessity"

Regarding sense of achievement, over 80% (81.81%) of students felt that having their outstanding assignments or works selected as teaching cases for subsequent courses would bring a sense of achievement (see Fig. 6). Meanwhile, over 60% (61.36%) were willing to publicly display their AI course works on the course platform, indicating a generally positive willingness to participate.

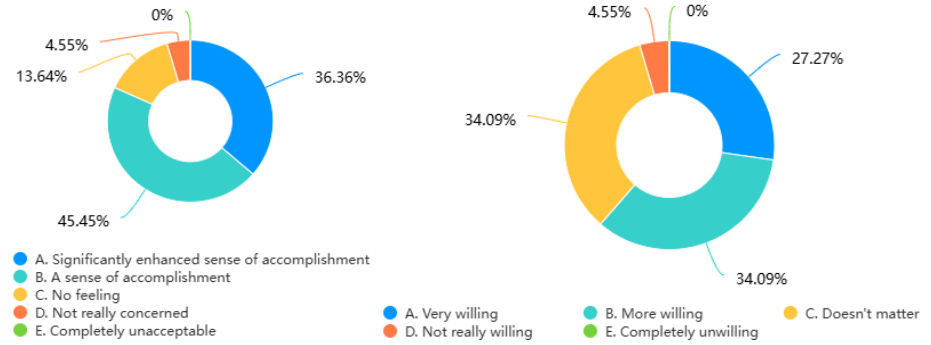


Fig 6. Student Achievement from Work Reused as Future Course Cases" and "Student Willingness for Public Work Display on Course Platform"

For the AI general education course, the top factor boosting student learning motivation is incorporating learning outcomes into course credits (59.09%), followed by selecting outstanding works as publicly displayed teaching cases (45.45%). Participating in assessment design and teamwork on real-world AI application projects both stand at 43.18%, while linking course content to competitions and innovation/entrepreneurship needs reaches 40.91%, and long-term recognition and display channels for learning outcomes accounts for 29.55%. These findings highlight the value of tangible rewards, public recognition and practical project experience.

5. Conclusion and Future Directions

5.1 Main Conclusions

This study builds an AI literacy curriculum for non-computer science undergraduates. The teaching practice show its content framework and pedagogical design effectively improve students' learning outcomes and comprehensive competencies. Process data reveal positive correlations between classroom interaction and team project quality, and between AI output evaluation quality and final exam scores. Students highly accept innovative assessments including team projects and self-designed questions. Learning motivation surveys confirm students prefer application- and safety-focused AI literacy content, validating the course design's rationality.

5.2 Limitations and Future Directions

This study is limited by a small sample size and lack of a control group. Future research should conduct multi-class and cross-institutional comparative experiments. Moreover, amid the rapid evolution of AI, further investigation is needed on dynamic course content updates, challenges from emerging technologies like Agentic AI to the literacy framework, and longitudinal tracking of students' long-term AI literacy retention. The international academic community has made initial responses, with relevant studies proposing a macro perspective shifting from mere technology integration to broader educational ecosystem transformation [13]. We believe that as more universities join the exploration of AI general education, a locally adapted and scalable AI literacy curriculum system will gradually mature.

References

[1]ManpowerGroup. *UK Talent Shortage*. 2026,
[2]ManpowerGroup. *Global Talent Shortage Reaches Turning Point as AI Skills Claim Top Spot*. 2026,
[3]Labor, U.S. Department of. *The U.S. Department of Labor's Artificial Intelligence Literacy Framework*. 2026,
[4]Long, Duri, et al., *What is AI Literacy? Competencies and Design Considerations*, in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 2020, Association for Computing Machinery: Honolulu, HI, USA. p. 1–16.
[5]Ng, Davy Tsz Kit, et al., *Conceptualizing AI literacy: An exploratory review.* Computers and Education: Artificial Intelligence, 2021. **2**: p. 100041.
[6]Online, People's Daily. *MOE Launches AI-Empowered Education Initiative; 85 'Golden Courses' from 64 Top Universities Go Live*. 2024,
[7]Wang, Yusu, *AI Empowerment and Interactive Digital Teaching via Rain Classroom: Innovative Practice of Foreign Language Teaching Reform in Higher Vocational Education.* The Educational Review, USA, 2026. **11**(2).
[8]University, Tsinghua. *Explore the Unknown, Touch the Future: This Course Brings You One Step Closer to AI*. 2024,
[9]university, Tsinghua. *Tsinghua University Artificial Intelligence Capability Enhancement Program Spring Semester 2025 Registration Opens*. 2025,
[10]University, Fudan. *Fudan University Launches 61 AI Courses for All Students in New Semester*. 2024,
[11]Net, Guangming. *Fudan University Develops Over 100 AI Mega-Courses to Transform Talent Cultivation Models as AI Enters Classrooms Across All Disciplines*. 2025,
[12]CAAI. *Stratified and Categorized, Progressive Enhancement: White Paper on AI General Education Curriculum System for Institutions of Higher Education*. 2025,
[13]Kremantzis, M, et al., *Moving Education Ecosystems Beyond the Artificial Intelligence Hype.* Intelligent Technologies in Education, 2026. **1**(1): p. 1-11.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)

© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261118

AI Innovative Practical Teaching Model for Interdisciplinary Postgraduates Integrated with OBE Concept

Jiawei Qiu, Jianghui Liu[1], Yulin Zheng and Chuyi Zou
Guangdong University of Foreign Studies, Guangzhou, China

Abstract. Traditional postgraduate education faces challenges in personalized guidance, dynamic competency tracking, and interdisciplinary practical training. This study proposed an AI-driven practical teaching model for interdisciplinary postgraduates based on the Outcomes-Based Education (OBE) concept. We adopted AI-assisted English learning as a representative application scenario and constructed a closed-loop teaching framework that integrates multi-agent collaboration, Large Language Models (LLM), and Retrieval-Augmented Generation (RAG). The framework establishes a multidimensional learner state model and a learning benefit function to support dynamic learner modeling, personalized task allocation, and intelligent feedback. We conducted a ten-week classroom teaching experiment involving interdisciplinary postgraduate students. The experiment compared three modes : traditional instruction, AI-agent-supported learning, and the integrated RAG–LLM teaching framework. The experimental results show that the proposed framework improves learning-path stability, and promote balanced development of vocabulary, grammar, listening, and speaking abilities. It also enhances the effectiveness and interpretability of feedback. The integrated RAG–LLM architecture facilitates the transition from passive learning to active learning. It also strengthens learner autonomy and strategic decision-making. This study provides an OBE-oriented AI teaching framework that integrate multi-agent collaboration with LLM and RAG technologies. The framework offers practical guidance for intelligent education and interdisciplinary postgraduate training.

Keywords. Smart education, Adaptive learning, Educational governance, Outcomes-Based Education (OBE), Retrieval Augmented generation (RAG), Large Language Model (LLM)

1. Introduction

With the rapid development of global technology and the wide application of artificial intelligence, higher education, especially postgraduate education, is facing unprecedented challenges and opportunities. The integration of artificial intelligence has made interdisciplinary research and application an inevitable trend. The number of interdisciplinary postgraduate students is increasing continuously. However, the

[1] Corresponding Author: Jianghui Liu, Center for Contemporary Education Technology, Guangdong University of Foreign Studies, Guangzhou, China; Email: 247031690@qq.com.

Supported by the Postgraduate Education Innovation Program Funding Project of Guangdong Province, China in 2025 (NO. 2025JGXM_065). Supported by the Undergraduate Innovation Training Project of Guangdong University of Foreign Studies in 2026.

traditional postgraduate education model still faces significant challenges in interdisciplinary integration, practical competency development, and personalized instruction. These limitations are particularly evident among interdisciplinary postgraduate students, who often possess diverse disciplinary backgrounds, varying levels of professional maturity, and highly individualized learning objectives. A standardized teaching approach is frequently unable to accommodate such differences, resulting in reduced learning efficiency and limited support for individual learning trajectories. Furthermore, postgraduate learners often need to balance academic research, professional development, and coursework simultaneously. This situation may increase the cognitive burden and reduce learning continuity. [1]

Recent advances in generative AI, represented by systems such as ChatGPT, DeepSeek and other large language models (LLMs), have significantly expanded the application of AI in education. AI technologies are increasingly used to support personalized learning, intelligent tutoring, automated assessment, and adaptive feedback, thereby creating new opportunities for improving teaching effectiveness and learner engagement.[2]

Therefore, integrating the OBE concept with AI technologies may provide an effective pathway for addressing these challenges. Compared with a conventional "one-size-fits-all" teaching models, an AI-driven adaptive learning systems can dynamically adjust learning tasks, generate personalized feedback, and continuously track a learner's progress. Such a mechanism helps reduce cognitive friction, improve learning efficiency, and enhance learner confidence by providing individualized support throughout the learning process. For interdisciplinary postgraduate students, this dynamic adaptation not only facilitates competency development but also accelerates the integration of disciplinary knowledge and practical skills.

This approach is also a key measure for cultivating high-quality talents for the future society [3]. By clarifying learning objectives, optimizing the curriculum design, innovating teaching methods, and establishing a diversified evaluation system, this model can effectively improve a postgraduate student's interdisciplinary integration ability and AI application ability. It also may provide theoretical support and practical experience for the further promotion of the OBE concept in China [4].

2. Research Status

In recent years, domestic and foreign research on the integration of AI and education has gradually deepened. Wang and Siau stated that AI can construct personalized learning paths and automated assessments to improve postgraduate learning efficiency. They also pointed out that generative AI such as ChatGPT has great potential in adaptive learning and language comprehension [5]. The U.S. Department of Education released the report Artificial Intelligence and the Future of Teaching and Learning: Insights and Recommendations in 2023. This report emphasized AI applications in adaptive learning, teaching optimization, teacher training and formative assessment. It advocated a human-centered teaching model and proposed solutions for context awareness, data security and model interpretability [6]. Some American universities have built AI-based adaptive learning platforms that provide customized learning resources and feedback based on a student's learning progress and performance. Meanwhile, the New Engineering Education Transformation (NEET) program at MIT adopts a problem-driven interdisciplinary teaching mode that effectively integrates scientific research and

teaching to promote educational innovation. Vaishya et al. studied the auxiliary value of AI in postgraduate examinations. The results showed that generative AI may improve problem-solving efficiency but lacks sufficient logical reasoning ability. Chang et al. adopted a hybrid mode combining AI tutors and human instructors. This mode helps strengthen postgraduates' career competence and explores the potential of human-machine collaboration. The Olin Triangle curriculum system and the thread settings of MIT-NEET reflect the application of the OBE concept in interdisciplinary teaching. Scholars from the American Society for Engineering Education (ASEE) have explored the implementation of the OBE concept in postgraduate curriculum design, teaching practice and evaluation [7][8].

3. Framework of OBE-Based AI Innovative Practical Teaching Model for Interdisciplinary Postgraduates

The AI practical teaching model for interdisciplinary postgraduates adopts an online and offline collaborative teaching mode. It aims to comprehensively improve students' AI practical ability and interdisciplinary innovation ability through various platforms and technical means. As shown in Figure 1, this model mainly consists of the following parts:

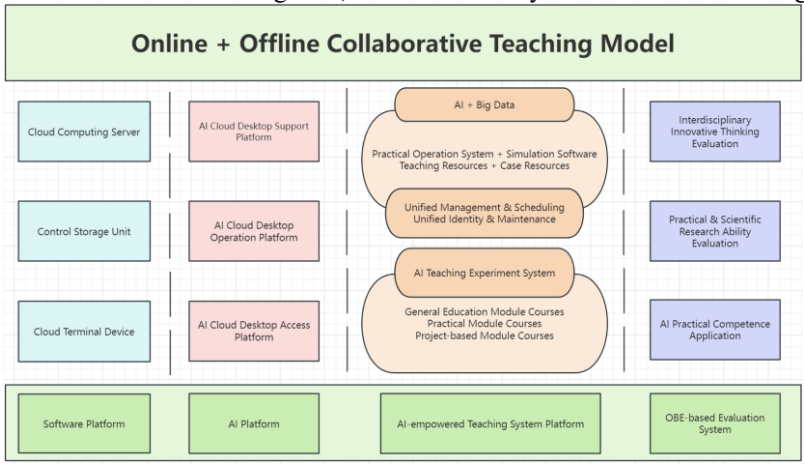


Figure 1. OBE-Based AI Practical Teaching Model for Interdisciplinary Postgraduates

Based on the above model, this study takes AI-assisted English learning as an example to illustrate the application of the teaching model in interdisciplinary postgraduate education. Unlike conventional language-learning systems, the proposed framework is designed to support the integration of disciplinary knowledge and language learning simultaneously. For interdisciplinary postgraduate students, English learning is closely associated with reading academic literature, understanding professional concepts, participating in international collaboration, and communicating research findings. English learning is selected as a representative implementation scenario because it naturally supports interdisciplinary knowledge acquisition, academic communication, and international collaboration among postgraduate students. Therefore, language learning serves not only as a linguistic training process but also as an important channel for acquiring domain-specific knowledge.

Unlike general English instruction, the proposed framework focuses on English for Academic Purposes (EAP) and English for Specific Purposes (ESP), where language learning is closely integrated with disciplinary knowledge acquisition. Accordingly, English learning activities are designed around authentic academic and professional contexts rather than isolated linguistic exercises.

The framework constructs an intelligent teaching framework that integrates learner modeling, agent collaboration, LLM+RAG driving, and dynamic optimization. This framework transforms the learning process from static practice into a closed-loop system with personalization, dynamic feedback, and optimized decision-making. Therefore, improving language competence through adaptive learning support can directly contribute to postgraduate academic development and professional growth. The framework has the potential to improve both learning effectiveness and the learning experience.

To further support interdisciplinary competency development, the proposed framework incorporates domain-specific knowledge into both conversational interactions and the RAG knowledge base. Learning materials can be dynamically generated from different disciplinary contexts, such as engineering, business management, data analytics, artificial intelligence, and other professional fields. In this way, learners simultaneously develop language competence and disciplinary expertise through a unified learning process.

For example, learners from different disciplinary backgrounds receive different learning resources through the RAG mechanism. Computer science students may retrieve recent AI research papers, software documentation, and algorithm descriptions to improve academic reading and technical writing skills. Business students may study international business cases, financial reports, and management literature to strengthen professional communication and presentation abilities. Engineering students may learn through technical reports, engineering standards, and project documentation. Although English learning serves as the implementation scenario, the retrieved knowledge, learning tasks, and expected competencies are dynamically adapted according to each learner's disciplinary backgrounds.

In order to better understand students' learning status, this study models the learner state. The English ability of a learner is described by a multi-dimensional vector. This mainly includes four dimensions: vocabulary, grammar, listening, and speaking. In the current framework, these dimensions are selected as foundational indicators for modeling learner progress and evaluating the effectiveness of adaptive learning interventions. Although they provide an effective basis for tracking language development, they primarily represent operational language mechanics and do not fully capture the higher-order academic and professional competencies required for interdisciplinary postgraduate education. Therefore, the present study focuses on establishing a reliable adaptive learning baseline, while more advanced competency dimensions will be incorporated in future extensions of the framework.

$$L_t = [\text{vocabulary}_t, \text{grammar}_t, \text{listening}_t, \text{speaking}_t] \quad (1)$$

Here, $\text{vocabulary}_t, \text{grammar}_t, \text{listening}_t, \text{speaking}_t \in [0, 1)$ represent the learner's ability levels in each dimension at time t. A larger value indicates a higher ability level

The learner's ability changes dynamically during the learning process. The variable $E_t$is used to represent the learner's practice performance vector at time t. It can include information such as accuracy, time used, and difficulty coefficient. The parameter α is

used to represent the learning rate, where $0 < \alpha \leq 1$, which controls the step size of the ability update.

Here, $f(\cdot)$ denotes the feedback function. It generates an adjustment value based on learning performance through the agent. The first term, $\beta_1 \cdot (E_t - L_t)$, represents the ability increment caused by the performance gap. A larger gap between the current ability and the target performance leads to a greater increment from this term. The second term, $\beta_2 \cdot S_t$, represents the ability increment caused by the learning state. A better learning state, such as high engagement and reasonable strategies, leads to a greater increment from this term.

$$f(E_t, S_t) = \beta_1 \cdot (E_t - L_t) + \beta_2 \cdot S_t \quad (2)$$

The variable S_t represents the behavioral preference vector. It includes risk preference, which refers to a preference for high-difficulty challenges. It also includes strategy preference, preference for intensive practice or extensive practice, engagement, learning investment level, and time preference, preference for short-time high-frequency or long-time low-frequency learning.

$$S_t = [risk_t, strategy_t, engagement_t, time_pref_t] \quad (3)$$

Based on the above analysis, the dynamic update formula can be expressed as:

$$L_{\{t+1\}} = L_t + \alpha \cdot f(E_t, S_t) \quad (4)$$

Through this dynamic modeling, the system can capture changes in a learner's abilities across all dimensions in real time. It provides a foundation for personalized practice. However, the system not only needs to track the learning status in real time. It also requires a circular teaching process to support learners in interdisciplinary learning.

At the current stage, the framework primarily supports foundational language learning tasks. However, the multi-agent architecture provides the scalability required for future expansion toward advanced academic and professional competency assessment. To ensure the reliability and educational appropriateness of AI-generated feedback, a verification and validation layer is incorporated into the proposed framework before information is delivered to learners. This layer functions as a quality-control mechanism that automatically evaluates generated responses from three perspectives: semantic consistency with retrieved knowledge, instructional relevance to learning objectives, and response confidence. Outputs that fail to satisfy predefined quality thresholds are either regenerated or flagged for additional review. Through this process, potential risks such as hallucinated content, retrieval inconsistencies, and knowledge conflicts can be effectively mitigated before reaching learners. To enhance the semantic understanding, reasoning, and interpretation capabilities of the system, a Large Language Model (LLM) is introduced. It is combined with the Retrieval-Augmented Generation (RAG) mechanism. This combination realizes knowledge-driven intelligent feedback and teaching support. A Large Language Model can perform semantic analysis, logical reasoning, and explanatory feedback on a student's answers. It outputs detailed explanations, learning suggestions, and strategic hints in natural language. For example: "Grammatical errors in this exercise mainly focus on tense usage. Please prioritize reviewing the differences between the past tense and the present perfect tense." RAG mechanism can retrieve relevant theories, typical examples, and cases from the teaching knowledge base. It improves the accuracy and interpretability of LLM-based feedback. For example: When a student makes a mistake in oral English, the RAG mechanism retrieves relevant pronunciation rules and sample sentences. Then the LLM generates targeted practice suggestions.

From the perspective of interdisciplinary learning, the integration of the LLM and the RAG mechanism creates a cross-disciplinary cognitive synergy effect. Traditional learning approaches often separate language acquisition from professional knowledge learning, forcing learners to repeatedly switch between different learning contexts. Such context switching may introduce cognitive downtime and reduce overall learning efficiency. In contrast, the proposed framework enables synchronous processing of disciplinary knowledge and language competencies within the same learning activity. By embedding professional content directly into learning tasks and feedback processes, learners continuously reinforce both knowledge domains, thereby shortening knowledge acquisition cycles and improving learning efficiency.

The combination of the LLM and the RAG mechanism demonstrates the potential to achieve interpretable feedback and personalized guidance. It supports multi-round dialogues. It meets learners' needs for personalized Q&A and in-depth learning. It ensures that feedback is both accurate and actionable. This study aims to replace fixed and passive instruction with dynamic algorithmic decision-making. It helps students shift from passively completing exercises to autonomous analysis and dynamic thinking. The goal is to improve students' AI practical ability, learner autonomy and self-regulation capability, and interdisciplinary literacy. In addition, the multi-agent collaboration mechanism dynamically adjusts task allocation, learning objectives, and feedback strategies according to learners' disciplinary backgrounds. Rather than providing identical learning paths for all learners, different agents coordinate discipline-specific resource retrieval, language assessment, and adaptive feedback. This coordination supports interdisciplinary competency development together with language proficiency improvement. Therefore, this study further optimizes the dynamic learning and algorithm framework. It adds an exercise allocation optimization model. It also improves the intelligent feedback and update formulas.

The exercise allocation model assigns the most suitable exercises to learners within a limited learning time. It aims to achieve maximized learning benefits. This study defines the learning benefit function as R_t. The variable L_t^i represents the learner's ability level in dimension i at time t, such as vocabulary, grammar, listening, and speaking. The variable w_i is the weight coefficient of each ability dimension, satisfying $\sum_i w_i = 1$. It can be dynamically adjusted according to teaching objectives. The function $g(\cdot)$ is a mapping function that maps ability levels to learning benefits. It is monotonically increasing and features diminishing marginal returns.

$$R_t = \sum_i w_i \cdot g(L_t^i) \tag{5}$$

In the agent feedback and update formula, the system generates feedback D_t based on the question Q_t and the learner state L_t. It determines the next exercise and updates the behavioral preference S_t simultaneously. The variable D_t represents the feedback content generated by the agent. It includes the correct answer, the error cause, the knowledge explanation, learning suggestions, and other details. The function $h(Q_t, L_t)$ is the feedback generation function. It generates personalized feedback by combining the current question and the learner state. The variable $Exercise_{t+1}$ denotes the next recommended exercise. It is determined by the learning benefit maximization model. The variable ΔS_t represents the change in behavioral preference. It is determined by the learning performance E_t and current ability L_t, and is used to update the behavioral preference S.

$$\begin{cases} D_t = h(Q_t, L_t) \\ Exercise_{t+1} = argmaxR_t \\ \Delta S_t = \phi(L_t, E_t) \end{cases} \quad (6)$$

This study further derives the updating formula for behavioral preferences. The function $\phi(L_t, E_t)$ stands for the behavioral preference updating function. The parameter η reflects the influence of learning experience on preferences and serves as the updating step size to control the changing speed of behavioral preferences.

$$S_{t+1} = S_t + \eta \cdot \Delta S_t = S_t + \eta \cdot \phi(L_t, E_t) \quad (7)$$

4. Conclusion

To facilitate practical deployment, future implementations should incorporate continuous verification, human supervision, and iterative optimization to improve the reliability and educational effectiveness of AI-generated feedback. This study demonstrates the effectiveness of the proposed AI-assisted teaching framework through classroom experimentation. The experimental results indicate that the proposed framework improves learning strategy stability, promotes balanced competency development, and enhances learner autonomy.

Overall, the AI innovative practical teaching model integrated with the OBE concept offers a practical and scalable solution for interdisciplinary postgraduate education. It potentially optimizes learning processes, improves learning efficiency, and strengthens learner autonomy and self-regulation capabilities in interdisciplinary postgraduate education. Unlike conventional English-learning systems that primarily improve general language proficiency, the proposed framework integrates discipline-specific knowledge retrieval with adaptive language learning through RAG and multi-agent collaboration. Consequently, language learning functions as a carrier for interdisciplinary competency development. This approach enables postgraduate students to improve both academic communication skills and professional knowledge simultaneously.

References

[1] Xue Y, Wang Y. Artificial Intelligence for Education and Teaching[J]. Wireless Communications and Mobile Computing, 2022.
[2] Sun L L, Balinas E. Integrating Outcome-Based Education (OBE) in Hybrid English Teaching Instruction for Higher Vocational Institutions[J]. 2024.
[3] Mahrishi M, Abbas A, Hosseini S, et al. A Bibliometric Analysis of the Global Trends of Outcome-Based Education (OBE) in Engineering Education[J]. 2024.
[4] Spady W. Outcome-based Education: Critical Issues and Answers[M]. Arlington, VA: American Association of School Administrators, 1994.
[5] Mufanti R, Carter D, England N. Outcomes-based education in Indonesian higher education[J]. 2024.
[6] Hao F, Zhao X. Design Framework and Practice of STEM Interdisciplinary project-based Teaching Activities Based on OBE Concept[J]. Journal of Educational Theory and Management, 2024, 8(4):41-45.
[7] Su R, Chaojie, et al. Reform and Exploration of the Trans Professional Experimental Teaching Model of Economics and Management based on the Concept of OBE——Take the Course of Enterprise Behavior Simulation as an Example[C]//2018.
[8] Huang H. Interdisciplinary Collaboration in Educational Informatization Innovation: Importance and Practice[J]. Research and Advances in Education, 2025, 4(1):38-46.DOI:10.56397/rae.2025.01.05.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261119

Predictive Validity of Online Learning Indicators for Collaborative Higher-Order Learning: Evidence from Learning Platform Data

Zixuan HUANG[a], Chunqiong WU[1,a], and Jianhong SUN[a]
[a] *School of Economics and Management, Yango University, Fuzhou, China*
ORCID ID: Chunqiong WU https://orcid.org/0009-0003-8621-9029

Abstract. Digital learning platforms generate extensive behavioral traces, but the predictive validity of individual indicators remains uncertain, especially for collaborative higher-order learning. Using anonymized LMS data from 177 undergraduates aggregated into 31 project groups, this study examined seven online learning indicators through descriptive statistics, correlation, regression, and sensitivity analyses. Chapter task completion showed an extreme ceiling effect, while attendance and interactive activity were also concentrated at the upper end of their scales. Chapter access frequency maintained the most stable positive association with final project performance, while video viewing duration added explanatory value when modeled jointly with access frequency. The findings show that digital traces should not be treated as equivalent measures of engagement and highlight the importance of measurement validity, outcome alignment, and group-level analysis in learning analytics.

Keywords. learning analytics; online learning indicators; digital learning traces; collaborative higher-order learning; educational data mining; higher education

1. Introduction

Learning management systems (LMS) routinely record task completion, resource access, video viewing, assessments, attendance, and participation. These digital traces create opportunities for learning analytics, but different indicators capture different aspects of learner behavior and should not be treated as equivalent measures of learning.

This distinction is particularly important for collaborative higher-order learning, which requires knowledge integration, analytical reasoning, problem solving, and communication. Because such outcomes are often demonstrated through group projects, directly linking individual behavioral records to a shared project score can create a mismatch in the unit of analysis.

---

[1] Corresponding Author: Chunqiong WU, Email: wu_chunqiong@foxmail.com, Tel: +86 13859025482
The authors acknowledge support from the 2023 Ministry of Education Humanities and Social Sciences General Project "From Behavioral Patterns to Decision Models: The Influence of Information-Seeking Modes on the Communication Effectiveness of Data Journalism in Liquid Scenarios" (Grant No. 23YJA840020), and the 2024 Fujian Provincial Science and Technology Plan Project: Mechanisms and Pathways of Digital Transformation of Agricultural Industry Chains Driven by New E-commerce (Grant No. 2024R0071).

Previous learning analytics studies demonstrate the predictive potential of digital traces, yet three challenges remain: indicators differ in educational meaning, composite course grades may overlap structurally with predictors, and individual-level records may not align with collaboratively produced outcomes.

This study addresses three research questions:

RQ1: Do different categories of online learning indicators demonstrate different levels of predictive validity for collaborative higher-order learning outcomes?

RQ2: Which online learning indicators provide meaningful explanatory value beyond routine assessment and participation measures?

RQ3: How do measurement characteristics, such as restricted variance and ceiling effects, influence the predictive usefulness of digital learning indicators?

To answer these questions, this study analyzes anonymized LMS data from 177 undergraduate students enrolled in a Big Data Analysis and Applications course. Individual learning behaviors were aggregated into 31 project groups and related to final project performance.

This study contributes by comparing multiple categories of platform indicators, while considering their measurement properties and aligning the analytical unit with the group level at which the learning outcome was produced.

The overall research framework is presented in Figure 1.

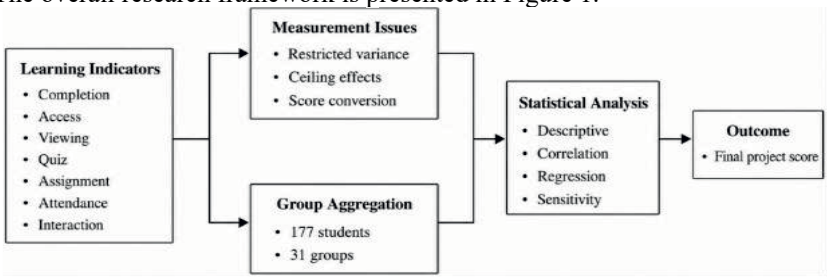


Figure 1. Research Framework

2. Related Work

2.1. Learning Analytics and Online Learning Indicators

Learning analytics (LA) refers to the collection, measurement, analysis, and reporting of learner-related data to understand and support learning processes [1]. LMS records have been used to identify behavioral patterns, predict academic outcomes and risks, and support interventions [2,3].

However, digital traces are indirect evidence rather than direct measurements of learning. Engagement includes behavioral, cognitive, and emotional dimensions [4], whereas platforms mainly capture observable actions such as accessing resources, completing tasks, or participating in activities.

Accordingly, the educational meaning of trace-based indicators requires theoretical justification and empirical validation [5]. Recent review evidence also identifies validity, generalizability, and temporal segmentation as key challenges in interpreting online trace data [6].

2.2. Predictive Use of Digital Traces in Learning Analytics

Educational data mining and machine learning have been widely used to predict examination performance and academic achievement [7,8]. Yet predictive accuracy alone does not establish the educational meaning of individual indicators.

Two methodological issues are especially relevant here. First, final grades often contain assignments, quizzes, attendance, or participation that are also used as predictors, creating criterion overlap. Second, collaborative projects produce shared outcomes, so treating individual behavioral records as independent observations can create pseudo-replication.

2.3. Measurement Challenges of Platform Indicators for Higher-Order Learning

Higher-order learning involves analysis, evaluation, knowledge application, and creation [9] and is often demonstrated through authentic projects. Many LMS indicators, however, were designed for course management rather than learning diagnosis.

When completion, attendance, or participation measures cluster near their maximum, ceiling effects reduce their ability to distinguish learners. Activity quantity also differs from learning quality: frequent access may indicate continued engagement, but it cannot reveal the cognitive purpose of each action.

Learning analytics should therefore consider measurement validity, context, and interpretability rather than predictive performance alone [10].

2.4. Collaborative Higher-Order Learning Assessment

Collaborative projects require analysis, knowledge transfer, communication, and coordination. Because LMS traces are generated at the individual level while project scores are shared, the analytical approach should align with the level at which the outcome is produced.

Collaborative learning analytics likewise emphasizes attention to group-level patterns, interaction structures, the interplay between individual and group levels, and temporality [11,12].

Accordingly, this study uses collaborative project performance as the higher-order learning outcome and aggregates individual LMS indicators to the project-group level before analysis.

3. Methodology

3.1. Research Context and Participants

This study used previously collected LMS data from an undergraduate Big Data Analysis and Applications course combining online activities, classroom instruction, and a group-based final project. The course aimed to develop data-processing, analytical, and applied problem-solving abilities.

Three course sections followed the same syllabus, learning activities, assessment criteria, and project requirements. LMS records included resource access, chapter tasks, online assessments, assignments, attendance, and interactive activities.

The dataset included 177 students. Students formed project groups voluntarily, largely on the basis of existing peer relationships and social preferences rather than researcher-assigned membership. The final project required problem formulation, data preparation, analytical modeling, interpretation, and reporting. Because each project received a shared group score, the 31 project groups served as the analytical units.

This design aligned the unit of analysis with the level at which the outcome was produced and avoided treating students sharing the same project score as independent observations.

3.2. Data Sources and Measures

Seven LMS indicators were analyzed and categorized as completion/compliance, content engagement, assessment performance, and participation. Table 1 summarizes their operational definitions.

Table 1. Definitions and Measurement of Study Variables

Category	Variable	Description
Completion and compliance	Chapter task completion	Platform-recorded completion level of assigned chapter tasks
Content engagement	Chapter access frequency	Number of times students accessed chapter learning resources
Content engagement	Video viewing duration	Total recorded duration of online video viewing
Assessment performance	Chapter quiz performance	Performance in online chapter quizzes
Assessment performance	Assignment performance	Performance in course assignments
Completion and compliance	Attendance score	Recorded participation in scheduled learning activities
Participation	Interactive activity score	Points obtained from online interactions such as discussions, questionnaires, and related activities
Learning outcome	Final project score	Group-level score of the comprehensive data analysis project

Chapter access frequency was defined as the number of recorded entries into chapter-level learning resources, including repeated visits to instructional materials.

The final project score served as the dependent variable. Because it was assigned at the group level, individual LMS indicators were aggregated by project membership before analysis.

Overall course grade and daily comprehensive score were excluded because they incorporated several predictor variables and would therefore introduce structural overlap between predictors and the outcome.

3.3. Data Preprocessing

The original dataset contained individual-level LMS records and group-level final project scores. Records were matched using anonymous student identifiers and checked for missing values, duplicates, and inconsistent group assignments; 177 valid student records remained.

For each project group, the seven student-level LMS indicators were aggregated using the arithmetic mean across members and matched with the shared final project score, yielding 31 group-level observations. Within-group behavioral dispersion was not modeled separately.

Distributional characteristics were examined before modeling. Substantially skewed behavioral variables were additionally transformed using $\ln(1+x)$ in sensitivity analyses.

3.4. Statistical Analysis

Descriptive statistics summarized the group-level variables. Pearson and Spearman correlations assessed linear and rank-order associations between each platform indicator and final project performance.

Multiple linear regression models evaluated indicator contributions, with final project score as the dependent variable. Three two-predictor category models were followed by a six-indicator integrated model; chapter task completion was excluded from the integrated model because 98.31% of students received the maximum score. HC3 robust standard errors were used for regression inference.

Sensitivity analyses excluded the singleton project group, applied $\ln(1+x)$ transformations to skewed variables, inspected standardized residuals, leverage, and

Cook's distance, used leave-one-group-out re-estimation, and compared raw behavioral measures with platform-converted scores.

All analyses used the project group as the analytical unit.

4. Results

The analyses used 31 project groups as independent observations, each represented by group-mean LMS indicators and one shared final project score.

4.1. Descriptive Statistics and Distributional Characteristics

Table 2 summarizes the group-level variables. Final project scores ranged from 45 to 90 (M = 70.35, SD = 10.84), indicating substantial variation across groups.

Table 2. Descriptive Statistics of the Group-Level Study Variables

Variable	n	Mean	SD	Minimum	Maximum	Skewness	Kurtosis
Chapter task completion	31	99.89	0.4	98.06	100	-3.94	15.76
Video viewing duration (minutes)	31	602.93	244.33	381.27	1650.33	2.94	11.04
Chapter access frequency	31	142.89	50.11	53	250	0.17	-0.65
Chapter quiz performance	31	77.65	10.33	52.97	96.39	-0.66	0.24
Assignment performance	31	83.91	7.78	55.4	92.99	-1.91	5.07
Attendance score	31	96	3.05	88.28	100	-0.98	1.02
Interactive activity score	31	90.68	16.15	17.87	100	-3.43	13.97
Final project score	31	70.35	10.84	45	90	-0.22	-0.35

Chapter task completion showed the strongest ceiling effect (M = 99.89, SD = 0.40); at the student level, 98.31% achieved the maximum and 99.44% scored above 90. Attendance and interactive activity were also concentrated at high values.

By contrast, chapter access frequency retained broad variation (M = 142.89, SD = 50.11; skewness = 0.17), while video viewing duration was more dispersed but positively skewed. Assessment-related indicators showed moderate variability.

4.2. Correlation Analysis

Table 3 reports Pearson and Spearman correlations between the seven platform indicators and final project performance.

Table 3. Correlations Between Online Learning Indicators and Final Project Score

Predictor	Pearson r	p	Spearman ρ	p
Chapter task completion	0.086	0.644	0.112	0.549
Video viewing duration	0.244	0.186	0.154	0.408
Chapter access frequency	0.451	0.011	0.53	0.002
Chapter quiz performance	0.155	0.406	0.148	0.427
Assignment performance	0.305	0.095	0.184	0.321
Attendance score	0.303	0.098	0.215	0.246
Interactive activity score	0.106	0.57	0.009	0.961

Chapter access frequency showed the strongest and most consistent association with project performance (Pearson r = 0.451, p = 0.011; Spearman ρ = 0.530, p = 0.002).

Video viewing duration was positively but not significantly associated with project performance (Pearson r = 0.244, p = 0.186; Spearman ρ = 0.154, p = 0.408). Assignment

performance and attendance had moderate positive Pearson correlations, but neither was statistically significant; the remaining indicators showed weaker associations.

Overall, continued resource access showed a stronger bivariate association with collaborative project performance than most completion- or participation-oriented indicators.

4.3. Multiple Regression Analysis

Table 4 summarizes four regression models. HC3 robust standard errors were used because of the limited number of group-level observations and possible heteroskedasticity.

Table 4. Regression Models Predicting Final Project Score

Model	Predictors	R^2	Adjusted R^2	Overall p-value	Main result
Model 1	Completion and compliance model	0.111	0.047	>0.05	Neither predictor was significant
Model 2	Content engagement model	0.257	0.204	0.01	Both predictors showed positive significant associations using HC3 standard errors
Model 3	Stage-based learning performance model	0.102	0.037	>0.05	Neither predictor was significant
Model 4	Integrated six-indicator model	0.397	0.246	0.056	Chapter access frequency remained significant

Given the small group-level sample, overall model significance should be interpreted cautiously; greater weight is placed on patterns that remain consistent across specifications.

The completion/compliance model was not significant ($R^2 = 0.111$, adjusted $R^2 = 0.047$). The content-engagement model performed best among the category models ($R^2 = 0.257$, adjusted $R^2 = 0.204$, $p = 0.010$): video viewing duration ($b = 0.0103$, $p = 0.022$) and chapter access frequency ($b = 0.0962$, $p = 0.027$) were both significant.

The stage-based learning performance model was not significant ($R^2 = 0.102$, adjusted $R^2 = 0.037$). In the integrated six-indicator model ($R^2 = 0.397$, adjusted $R^2 = 0.246$, overall $p = 0.056$), chapter access frequency remained significant ($b = 0.0963$, $p = 0.003$).

Across specifications, chapter access frequency was the most consistent predictor of collaborative project performance.

4.4. Sensitivity Analyses and Robustness Checks

Table 5 summarizes sensitivity analyses under alternative data-processing and model specifications.

Table 5. Ceiling Characteristics and Associations of Selected Platform Indicators with Project Performance

Indicator	Students at maximum score (%)	Students scoring at least 90 (%)	Pearson correlation with project performance
Chapter task completion	98.31	99.44	0.086
Attendance score	59.32	84.75	0.303
Interactive activity score	67.23	72.88	0.106
Converted chapter learning score	42.37	44.07	0.483
Raw chapter access frequency	Not applicable	Not applicable	0.451

Excluding the singleton group and log-transforming skewed variables did not materially change the main pattern. Some groups showed elevated leverage or Cook's

distance, but leave-one-group-out re-estimation preserved the positive chapter-access coefficient.

Raw chapter access frequency preserved uncapped behavioral variation, whereas the converted score placed 42.37% of students at the maximum; both measures were positively associated with project performance.

Together, these checks support the robustness of the association between chapter access frequency and collaborative project performance.

5. Discussion

The findings show that digital traces do not provide equivalent evidence of learning: predictive usefulness depends on the behavior captured, measurement characteristics, and alignment with the target outcome. Routine compliance measures were weakly informative, whereas repeated resource access was more consistently associated with project performance.

5.1. Limited Predictive Power of Routine Platform Metrics

Completion, attendance, and participation scores had limited predictive value for collaborative higher-order outcomes. Their usefulness for monitoring course requirements does not imply diagnostic value when ceiling effects compress variation.

Participation measures also record whether interaction occurs rather than its quality. Quiz and assignment performance showed more variation but mainly reflected structured achievement, whereas the project required knowledge integration, open-ended analysis, and coordination.

Learning analytics should therefore distinguish administrative monitoring from learning diagnosis; the availability of behavioral data alone does not guarantee meaningful evidence of learning.

5.2. Why Chapter Access Frequency Mattered

Chapter access frequency was the most stable predictor across correlation, regression, and robustness analyses, suggesting that repeated interaction with resources was more closely related to project-based learning processes than simple completion or participation.

Project work is iterative rather than linear. Students may return to instructional resources to review procedures, refine methodological choices, or resolve problems during analysis, making repeated access an observable trace of continued interaction with learning materials.

However, access frequency is not a direct measure of learning quality. Repeated visits may reflect productive review, uncertainty, or inefficient searching, and their meaning may also depend on when they occur in the project lifecycle (e.g., sustained access versus pre-deadline clustering). The finding therefore supports predictive usefulness rather than a causal mechanism.

The comparison with the platform-converted chapter learning score shows that assessment-oriented transformations can compress behavioral variation. Retaining raw access counts preserves variation that is lost once the converted score reaches its maximum, even though both measures are associated with performance.

5.3. Ceiling Effects as a Measurement Validity Challenge in Learning Analytics

Ceiling effects illustrate a measurement-validity problem: an indicator can accurately record compliance yet provide little predictive information once most learners reach the maximum.

Measures used for course management should therefore not automatically be interpreted as evidence of higher-order learning. More informative analytics may require process-oriented indicators that preserve meaningful behavioral variation.

5.4. Implications for Learning Analytics Design and Educational Decision-Making

Platform indicators should be interpreted according to their educational meaning rather than selected simply because they are easy to collect. Completion and attendance remain useful for course management, whereas process-oriented indicators may provide richer information for learning diagnosis.

Educational platforms should preserve raw behavioral records alongside converted assessment scores because transformations useful for grading can reduce the variation needed for predictive analysis.

Learning analytics should function as decision support rather than automated judgment. Digital traces cannot fully capture motivation, cognitive processes, or collaborative contributions and should be interpreted with contextual evidence and teacher judgment.

For collaborative learning, future systems should also consider within-group contribution patterns and interaction structures.

6. Conclusion and Future Research

This study demonstrates that online learning indicators represent different behavioral dimensions and differ in predictive usefulness. Completion- and participation-oriented measures showed limited explanatory value under ceiling effects, whereas chapter access frequency maintained the most stable association with collaborative project performance; video viewing duration added value when modeled jointly with access frequency.

The findings support indicator-specific interpretation, preservation of raw behavioral variation, and alignment between the analytical unit and the level at which the learning outcome is produced.

6.1. Theoretical and Practical Contributions

Theoretically, the study highlights the distinction between administrative monitoring indicators and diagnostic evidence of complex learning, and shows why measurement properties matter for predictive validity.

Methodologically, using an independently assessed group project reduces predictor-criterion overlap, while aggregating individual behaviors to the project-group level aligns predictors with a collaboratively produced outcome.

Practically, learning analytics systems should retain interpretable process data and support contextualized instructional decisions rather than treat platform scores as isolated measures of student ability.

6.2. Limitations and Future Research

This study has several limitations. First, it was conducted in a single course context with 31 project-group observations, which limits statistical power. Future research should validate the findings using larger datasets across courses, disciplines, and institutions.

Second, the observational design does not allow causal conclusions. The relationship between chapter access frequency and project performance may also reflect learner characteristics such as motivation, prior knowledge, or self-regulated learning ability. Future studies should incorporate longitudinal designs or additional learner-level variables.

Third, platform indicators capture only observable behaviors and cannot fully represent cognitive processes or collaborative contributions. Future research should integrate additional evidence, such as interaction data, revision behaviors, and qualitative learning processes, to develop richer representations of collaborative learning.

References

[1] Siemens G, Long P. Penetrating the fog: analytics in learning and education. EDUCAUSE Rev. 2011;46(5):30-32,34,36,38,40.
[2] Arizmendi CJ, Bernacki ML, Raković M, Plumley RD, Urban CJ, Panter AT, Greene JA. Predicting student outcomes using digital logs of learning behaviors: review, current standards, and suggestions for future work. Behav Res Methods. 2023;55:3026-3054. doi:10.3758/s13428-022-01939-9.
[3] Xiao W, Ji P, Hu J. A survey on educational data mining methods used for predicting students′ performance. Eng Rep. 2022;4(5):e12482. doi:10.1002/eng2.12482.
[4] Fredricks JA, Blumenfeld PC, Paris AH. School engagement: potential of the concept, state of the evidence. Rev Educ Res. 2004;74(1):59-109. doi:10.3102/00346543074001059.
[5] Fan Y, van der Graaf J, Lim L, Raković M, Singh S, Kilgour J, Moore J, Molenaar I, Bannert M, Gašević D. Towards investigating the validity of measurement of self-regulated learning based on trace data. Metacogn Learn. 2022;17:949-987. doi:10.1007/s11409-022-09291-1.
[6] Du J, Hew KF, Liu L. What can online traces tell us about students' self-regulated learning? A systematic review of online trace data analysis. Comput Educ. 2023;201:104828. doi:10.1016/j.compedu.2023.104828.
[7] Yağcı M. Educational data mining: prediction of students' academic performance using machine learning algorithms. Smart Learn Environ. 2022;9:11. doi:10.1186/s40561-022-00192-z.
[8] Zhang Y, Yun Y, An R, Cui J, Dai H, Shang X. Educational data mining techniques for student performance prediction: method review and comparison analysis. Front Psychol. 2021;12:698490. doi:10.3389/fpsyg.2021.698490.
[9] Anderson LW, Krathwohl DR, editors. A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives. New York: Longman; 2001.
[10] Scalise K, Wilson M, Gochyyev P. A taxonomy of critical dimensions at the intersection of learning analytics and educational measurement. Front Educ. 2021;6:656525. doi:10.3389/feduc.2021.656525.
[11] Martinez-Maldonado R, Gašević D, Echeverria V, Fernandez Nieto G, Swiecki Z, Buckingham Shum S. What do you mean by collaboration analytics? A conceptual model. J Learn Anal. 2021;8(1):126-153. doi:10.18608/jla.2021.7227.
[12] Wise AF, Rosé CP, Järvelä S. Editorial: Nine elements for robust collaborative learning analytics: a constructive collaborative critique. Int J Comput Support Collab Learn. 2023;18:1-9. doi:10.1007/s11412-023-09389-x.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)

© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261120

Research on the Influencing Mechanism of Higher Vocational Students' Entrepreneurial Intention Based on Three-Dimensional Driving Mediation Model

Xiaoqi Huang[1] and Shaozhong Zhang
Guangdong Polytechnic of Industry and Commerce, Guangzhou 510550, China

Abstract. Against the background of innovation-oriented higher vocational education, exploring the formation mechanism of college students' entrepreneurial competency is of great significance for improving the effectiveness of innovation and entrepreneurship education. Grounded on the Theory of Planned Behavior, Social Cognitive Theory and Entrepreneurial Ecosystem Theory, this paper constructs a theoretical framework labeled "three-dimensional drivers-competency mediator-intention output". Structural Equation Modeling (SEM) is adopted to test research hypotheses and analyze mediating effects based on 467 valid questionnaire samples. The results show that three systems including family support, school education and social environment exert significant positive impacts on students' individual comprehensive competency. Individual comprehensive competency generates a strong positive effect on entrepreneurial intention and cultivation outcomes. It exerts significant mediating effects across the three pathways, among which partial mediation is verified for family support and school education. For the social environment pathway, the indirect effect is significant while the direct effect is marginally non-significant, and the mediating effect is supported. Most effects of external driving factors are realized through the internal transformation of individual comprehensive competency, and school education produces the strongest driving effect. Questionnaire statistics indicate that students hold relatively high entrepreneurial interest, yet their actual post-graduation entrepreneurial willingness remains low. This study verifies that entrepreneurial intention is not the direct output of external conditions, but an outcome transformed via the internalization of individual competency. The findings can provide empirical references for higher vocational colleges to build a family-school-social collaborative entrepreneurship cultivation system.

Keywords. Higher vocational students; Entrepreneurial competency; Entrepreneurial intention; School education; Family support

1. Introduction

Against the backdrop of innovation-oriented vocational education development, seven ministries including the Ministry of Human Resources and Social Security have issued

[1] Corresponding Author: Xiaoqi Huang, 164794259@qq.com.

documents to improve the entrepreneurship education system of colleges and strengthen the cultivation of students' entrepreneurial awareness and capabilities [1]. As an indispensable component of higher education, higher vocational colleges have continuously increased the supply of innovation-entrepreneurship courses and competition activities. Nevertheless, a practical dilemma emerges characterized by high entrepreneurial interest yet low conversion rate: there exists an obvious "black box" between students' entrepreneurial willingness and actual entrepreneurial behavior [2]. Existing studies mainly concentrate on school-based innovation and entrepreneurship education reforms, while in-depth explorations on the internal transmission mechanism generated by the joint effect of multiple systems remain insufficient. Accordingly, this paper constructs a theoretical framework and conducts questionnaire-based empirical analysis to dissect the formation mechanism of entrepreneurial competency, so as to provide empirical evidence for improving the effectiveness of innovation and entrepreneurship education.

This study delivers both theoretical and practical values. Theoretically, most existing empirical researches related to vocational-college entrepreneurship focus on the effect of single variables, and empirical evidence concerning the mediating transmission mechanism of entrepreneurial competency among higher vocational student groups still needs to be supplemented [2]. This research enriches empirical achievements regarding the formation mechanism of entrepreneurial competency within vocational education. Practically, under the national policy orientation advocating multi-stakeholder collaborative talent cultivation [3], the research conclusions can offer references for constructing entrepreneurship cultivation systems and help promote the performance of innovation and entrepreneurship education in higher vocational colleges.

Most previous empirical entrepreneurship researches adopt a school-centered research perspective [4]. Although several studies recognize the vital values of family and social environments, their discussions are limited to qualitative analysis or separate tests of a certain type of external variables [5]. Few scholars integrate the three major systems of family, school and society into a unified structural equation model. Taking 467 valid questionnaire responses collected from higher vocational students in Guangdong Province as samples, this paper builds the theoretical model labeled "three-dimensional drivers-competency mediator-intention output" based on the Theory of Planned Behavior, Social Cognitive Theory and Entrepreneurial Ecosystem Theory. Structural Equation Modeling (SEM) is adopted to carry out reliability and validity tests, path analysis and Bootstrap-based mediating effect tests, so as to investigate the functional relationships among three external systems (family, school, society), individual comprehensive competency and entrepreneurial intention. Following the technical route of "theoretical construction-questionnaire investigation-data verification-conclusions and implications", this paper reveals the formation mechanism of entrepreneurial intention, provides empirical support for family-school-social collaborative entrepreneurship cultivation, and further fills the research gaps of existing literature.

2. Theoretical Basis and Research Hypotheses

With the continuous advancement of innovation and entrepreneurship education, scholars have widely recognized that the diversified environments of family, school and

society jointly shape students' entrepreneurial intention. Nevertheless, existing studies mostly examine the influence of a certain type of environmental factor in isolation, and integrated analysis of the complete transmission chain of "external environment-internal competency-entrepreneurial intention" still needs further deepening. To clarify how the three-dimensional external system of family-school-society acts on college students' entrepreneurial intention through individual internal competency, this chapter sorts out relevant classical theories and empirical literature, defines core research concepts, and accordingly puts forward a series of research hypotheses for this study.

2.1. Theoretical Basis

Proposed by Icek Ajzen, the Theory of Planned Behavior establishes the driving mechanism of three cognitive elements including attitude toward behavior, subjective norm and perceived behavioral control on behavioral intention, which provides a core theoretical framework for understanding the psychological generation path of individual entrepreneurial intention [6]. Krueger and Carsrud first introduced the Theory of Planned Behavior into entrepreneurship research. They verified that entrepreneurial behavior is essentially a planned behavior, and entrepreneurial intention can be effectively predicted by attitude and perceived behavioral control, laying an empirical foundation for the formation logic of entrepreneurial intention in this study [7]. Krueger compared the Entrepreneurial Event Model and the Theory of Planned Behavior, and found that the two models have equivalent predictive power for entrepreneurial intention. Among them, the concept of perceived behavioral control highly overlaps with perceived feasibility, which reinforces the core position of cognitive elements in the formation of entrepreneurial intention [8]. Bandura's Social Cognitive Theory proposes the reciprocal determinism of "environment-individual-behavior". It reveals that external environments do not exert effects on behavioral intention directly, but function through the mediating transmission of individual cognitive evaluation and self-efficacy, which offers theoretical evidence for the mechanism that external environments affect entrepreneurial intention via individual competency [9]. The Entrepreneurial Event Theory proposed by Shapero and Sokol further complements the environmental triggering mechanism. It holds that the initiation of entrepreneurial behavior depends on breaking individual inertia and individual judgment toward entrepreneurship. Perceived feasibility shares the same theoretical origin with perceived behavioral control in the Theory of Planned Behavior. Together they support the theoretical logic that external environments can be transformed into entrepreneurial intention only through individual cognitive internalization [10]. Bronfenbrenner's Social-Ecological Systems Theory investigates individual development within multi-layered nested environmental systems. Family belongs to the microsystem with direct interaction; school education falls into the mesosystem; social policies and entrepreneurial culture pertain to the exosystem and macrosystem. Interactions such as linkage, complementarity and offset exist among environments at different levels. It provides theoretical support of environmental stratification for constructing the "three-dimensional drivers" external input framework in this research [11]. From the perspective of human capital, Martin, McNally and Kay confirmed through meta-analysis that entrepreneurship education can significantly improve students' entrepreneurial knowledge, skills and entrepreneurial intention. As a vital approach for

human capital accumulation, school entrepreneurship education serves as a critical channel for external environments to be transformed into individual competency [12]. In terms of family environment, empirical research by Zhang Zeqing and Huang Zerui demonstrates that family capital significantly affects students' entrepreneurial intention through the chain mediating effect of social support and critical thinking [13]. Recent studies further indicate that family emotional support indirectly facilitates the formation of students' entrepreneurial intention by enhancing entrepreneurial self-efficacy and perceived behavioral control. Moreover, family support plays a significant moderating role in the transformation from entrepreneurial intention to actual behavior [14]. The above-mentioned theoretical and empirical findings jointly support the complete research framework of "three-dimensional drivers-competency mediator-intention output".

2.2. Definition of Core Concepts

In this paper, the three-dimensional driving system refers to an external cultivation system consisting of family support, school education and social environment. Corresponding to three categories of questionnaire items about family resources, school innovation and entrepreneurship supply, and social entrepreneurial atmosphere, it delivers external inputs for the formation of entrepreneurial competencies [15]. As the mediating variable, individual comprehensive competency covers entrepreneurial awareness, practical operation ability, frustration resistance, physical and mental literacy as well as academic literacy, measured by self-evaluation items in questionnaires. It acts as the internal mediator that converts external environmental factors into entrepreneurial intention [16]. As the outcome variable, entrepreneurial intention reflects students' subjective psychological tendencies including entrepreneurial interest, career preference and willingness to participate in entrepreneurship education. It is a predictive precursor of entrepreneurial behavior and measured by willingness-related questionnaire items.

2.3. Research Hypotheses

Based on the above theoretical framework, combined with existing research outcomes and practical observations, seven research hypotheses are proposed as follows:
H1: The family support system exerts a significant positive impact on college students' individual comprehensive competency.
H2: The school education system can significantly and positively predict college students' individual comprehensive competency.
H3: The social environment system has a significant positive effect on individual comprehensive competency.
H4: Individual comprehensive competency exerts a significant positive influence on entrepreneurial intention and cultivation outcomes.
H5: Individual comprehensive competency plays a mediating role between family support and entrepreneurial intention.
H6: Individual comprehensive competency plays a mediating role between school education and entrepreneurial intention.
H7: Individual comprehensive competency plays a mediating role between social environment and entrepreneurial intention.

3. Research Design

This section elaborates on the implementation of questionnaire survey, the selection basis of observed variables and the specification of the structural equation model. In line with the foregoing research hypotheses, this study takes the family support system, school education system and social environment system as exogenous latent variables, individual comprehensive competency as the mediating latent variable, and entrepreneurial intention and cultivation effect as the endogenous outcome latent variable. Multi-dimensional observed indicator data corresponding to latent variables are collected via questionnaires to construct the structural equation model for subsequent hypothesis verification.

3.1. Questionnaire Design and Sample Profile

A total of 467 valid questionnaires were collected in this survey. The respondents are mainly first-year and second-year higher vocational students in Guangdong Province. The questionnaire items cover six modules: individual characteristics, family background, school education, social environment, individual comprehensive competency, and entrepreneurial intention.

3.2. Variable Measurement

3.2.1. Exogenous Latent Variables

(1) Family Support System (ξ) This latent variable reflects family-level support for cultivating college students' entrepreneurial competency, with seven observed indicators: annual household income, father's educational attainment, mother's educational attainment, parental entrepreneurial experience, harmony of family atmosphere, father's entrepreneurial support, and mother's entrepreneurial support. Among them, the composite indicator of parental entrepreneurial support is calculated as the average score of the two items.
(2) School Education System (ξ) This latent variable represents the entrepreneurship-related support provided by higher vocational colleges, consisting of nine observed indicators: satisfaction with entrepreneurship courses, satisfaction with entrepreneurship instructors, richness of entrepreneurship activities, construction of entrepreneurial practical training bases, implementation of campus entrepreneurial policies, integration of professional and entrepreneurship education, integration of ideological and entrepreneurship education, participation frequency of entrepreneurship competitions, and institutional emphasis on entrepreneurship education.
(3) Social Environment System (ξ) This latent variable reflects the macro-level entrepreneurial institutional environment, public services and cultural atmosphere. It includes seven observed indicators: satisfaction with social credit system, friendliness of entrepreneurial environment, satisfaction with public entrepreneurial service system, completeness of relevant policies and regulations, soundness of institutional mechanisms, urban entrepreneurial atmosphere, and convenience of government administrative services.

3.2.2. Mediating Latent Variable

Individual Comprehensive Competency (η) This latent variable reflects the overall development level of college students and contains eight observed indicators: entrepreneurial awareness, practical operational ability, stress and frustration resistance, verbal communication ability, physical health status, mental health status, academic performance of major courses, humanistic literacy and clarity of learning objectives.

3.2.3. Endogenous Outcome Latent Variable

Entrepreneurial Intention and Cultivation Effect (η_2) This latent variable represents the final output of entrepreneurial competency cultivation, including four observed indicators: interest in self-employment, self-perceived suitability for entrepreneurship, willingness to receive entrepreneurship education, and career tendency to start a business after graduation.

All observed variables are measured by a 5-point Likert scale, with scores ranging from 1 to 5 corresponding to increasing degrees. Demographic variables are incorporated into the model as control variables.

3.3. Construction of Structural Equation Model

Structural Equation Modeling (SEM) is adopted in this study to verify the research hypotheses. The measurement model characterizes the correlations between latent variables and their corresponding observed indicators, while the structural model depicts the path relationships among latent variables.

The matrix expressions of the model are defined as follows:

Measurement Model: $X=\lambda x\xi+\delta; Y=\lambda y\eta+\varepsilon$

Structural Model: $\eta=\beta\eta+\Gamma\xi+\zeta$

Where $\xi=[\xi1,\xi2,\xi3]$ denotes the vector of exogenous latent variables, corresponding to the Family Support System, School Education System and Social Environment System respectively. $\eta=[\eta1, \eta2]$ T stands for the vector of endogenous latent variables, representing Individual Comprehensive Competency as well as Entrepreneurial Intention and Cultivation Effect. λx and λy refer to factor loading matrices. β is the path coefficient matrix among endogenous latent variables, and Γ represents the matrix of path coefficients from exogenous latent variables to endogenous latent variables. δ、ε、ζ represent measurement errors and structural residual terms respectively.

4. Empirical Analysis

This chapter conducts reliability-validity test, overall model fitness evaluation, path coefficient analysis and Bootstrap-based mediating effect test sequentially. Firstly, the reliability and validity of the measurement model are examined to ensure that observed variables can effectively reflect their corresponding latent variables. Secondly, the overall fitting performance of the structural equation model is assessed. Then, path coefficients corresponding to each research hypothesis are tested. Finally, mediating effect decomposition is performed to analyze the internal transmission mechanism through which external systems affect entrepreneurial intention via individual comprehensive competency.

4.1. Reliability and Validity Test

Cronbach's α coefficient and composite reliability (CR) are jointly adopted to test the scale reliability. As shown in Table 1, the scale achieves satisfactory reliability.

Table 1. Reliability Test Results of All Latent Variables

Latent Variable	Number of Observed Items	Cronbach's α	Composite Reliability (CR)	Reliability Evaluation
Family Support System	7	0.714	0.776	Good
School Education System	9	0.883	0.902	Excellent
Social Environment System	7	0.736	0.781	Good
Individual Comprehensive Competency	8	0.819	0.854	Excellent
Entrepreneurial Intention and Cultivation Effect	4	0.702	0.748	Good
Overall Scale	35	0.905	—	Excellent

The results of Confirmatory Factor Analysis (CFA) in Table 2 verify favorable discriminant validity of the measurement scale.

Table 2. Convergent Validity and Discriminant Validity of Latent Variables

Latent Variable	AVE	1	2	3	4	5
1. Family Support System	0.501	0.708				
2. School Education System	0.565	0.357	0.752			
3. Social Environment System	0.513	0.311	0.432	0.716		
4. Individual Comprehensive Competency	0.527	0.403	0.491	0.374	0.726	
5. Entrepreneurial Intention and Cultivation Effect	0.518	0.332	0.426	0.325	0.604	0.72

Note: The bold diagonal values represent square roots of AVE; the lower-triangle entries are correlation coefficients between latent variables.

4.2. Overall Model Fit

Table 3 presents the global fit indices of the structural equation model. The model fits well with sample data, which verifies the rationality of the theoretical model.

Table 3. Global Fit Indices of Structural Equation Model

Fit Indicator	χ2/df	GFI	AGFI	CFI	TLI	RMSEA	SRMR
Value	2.427	0.901	0.872	0.924	0.91	0.064	0.057
Threshold	<3	>0.9	>0.8	>0.9	>0.9	<0.08	<0.08
Evaluation	Good	Good	Acceptable	Excellent	Good	Good	Good

4.3. Path Coefficient Test

The maximum likelihood estimation method is applied to estimate the path coefficients of the model, and the results are summarized in Table 4.

Table 4. Summary of Standardized Path Coefficients and Hypothesis Test Results

Path Relationship	Standardized Path Coefficient β	Standard Error (SE)	Z-value	p-value	Test Result
ξ1 Family Support System→η1 Individual Comprehensive Competency	0.227	0.048	4.729	***	Supported
ξ2 School Education System→η1 Individual Comprehensive Competency	0.351	0.051	6.882	***	Supported
ξ3 Social Environment System→η1 Individual Comprehensive Competency	0.184	0.046	4.000	***	Supported
η1 Individual Comprehensive Competency →η2 Entrepreneurial Intention and Cultivation Effect	0.473	0.053	8.925	***	Supported
ξ1 Family Support System→η2 Entrepreneurial Intention and Cultivation Effect	0.116	0.045	2.578	0.010	Supported
ξ2 School Education System→η2 Entrepreneurial Intention and Cultivation Effect	0.172	0.05	3.440	***	Supported
ξ3 Social Environment System→η2 Entrepreneurial Intention and Cultivation Effect	0.089	0.047	1.894	0.058	Not Supported

The test results reveal that the school education system exerts the strongest positive influence on individual comprehensive competency (β=0.351, p<0.001), followed by the family support system (β=0.227, p<0.001), while the social environment system shows a relatively weaker effect (β=0.184, p<0.001). Moreover, individual comprehensive competency generates the largest predictive effect on entrepreneurial intention (β=0.473, p<0.001), which acts as the core determinant of entrepreneurial intention.

4.4. Mediation Effect Test

The bias-corrected Bootstrap method is utilized to examine the mediating effect of individual comprehensive competency, and the corresponding outcomes are displayed in Table 5. The 95% confidence intervals of indirect effects for three pathways do not contain zero, confirming that all three mediating pathways are statistically significant. For family support and school education, both direct and indirect effects are significant, supporting partial mediation. Regarding the social environment pathway, the indirect effect is significant, while its direct effect is marginally non-significant (P=0.058). Therefore, the mediating effect holds, yet we refrain from rigidly labelling it as full mediation given the marginal statistical result. Approximately 47%-49% of the total effect of each external driving factor on entrepreneurial intention is transmitted via individual comprehensive competency.

Table 5. Decomposition and Test Results of Mediation Effects

Pathway	Direct Effect	Indirect Effect	Total Effect	Proportion of Mediation Effect	95% Confidence Interval	Mediation Type
Family Support → Entrepreneurial Intention	0.116	0.107	0.223	47.98%	[0.121,0.332]	Partial Mediation
School Education → Entrepreneurial Intention	0.172	0.166	0.338	49.11%	[0.230,0.455]	Partial Mediation
Social Environment → Entrepreneurial Intention	0.089	0.087	0.176	49.43%	[0.081,0.278]	Mediation supported

5. Conclusions and Practical Implications

The empirical results indicate that family support, school education, and social environment all exert significant positive effects on students' individual comprehensive competency, among which school education presents the strongest driving effect. Individual comprehensive competency has a strong positive influence on entrepreneurial intention and plays significant mediating roles between external systems and entrepreneurial intention. Partial mediation holds for the pathways of family support and school education. As for the social environment pathway, the indirect effect is statistically significant whereas the direct effect is marginally non-significant, and the mediating effect is supported. Sample statistics reveal that higher vocational students possess relatively high entrepreneurial interest yet low actual entrepreneurial willingness, demonstrating that external conditions cannot directly generate entrepreneurial intention.

Accordingly, practical implications are proposed. First, higher vocational colleges should consolidate the dominant position of innovation and entrepreneurship education and focus on cultivating students' internal competencies such as entrepreneurial awareness and frustration resistance. Second, collaborative communication channels between families and colleges should be established to attach importance to the subtle educational influence of families. Through family-school-social collaboration, external resources can be effectively converted into students' internal entrepreneurial competency, so as to improve the practical effectiveness of innovation and entrepreneurship education in higher vocational colleges.

The samples of this study are mainly collected from higher vocational students in Guangdong Province, which brings limitations to sample coverage. Future research can expand the sample scope by including higher vocational colleges from other provinces to further deepen the research on the cultivation mechanism of entrepreneurial competency.

Acknowledgements.

This work was supported by the Teaching Innovation Team Project of Economy and Trade Innovation & Entrepreneurship, Guangdong Polytechnic of Industry and Commerce (2024); and the Horizontal Project "Research on the Innovation Mechanism of Industry-Education Integration for College Students" Employment and Entrepreneurship Education" of Guangdong Polytechnic of Industry and Commerce.

References

[1] Ministry of Human Resources and Social Security of China, et al. Opinions on Improving the Entrepreneurship Support System and Enhancing Entrepreneurship Quality [EB/OL].

[2] Wu QS, Hu H. Is there a "black box" between higher-vocational students' entrepreneurial intention and entrepreneurial behavior? Construction and test of a moderated multiple-mediation model. Vocational and Technical Education, 2023, 8: 57-63.

[3] Ministry of Education of China. Opinions on Deepening the Reform of Key Teaching Elements in Vocational Education [EB/OL].

[4] Xu B, Deng W B. Research on the Influencing Factors of Higher Vocational Entrepreneurship Education Based on Structural Equation Model—An Empirical Study by Entrepreneurial Ability and Integration

of Professional Education and Entrepreneurship Education. Journal of Vocational and Technical College, 2025, 1: 37-44.
[5] Zhang ZQ, Huang ZR. Influence of family capital on vocational-college undergraduates' entrepreneurial willingness: Chain mediating effect of social support and critical thinking. Psychological Monthly, 2024, 14: 72-75.
[6] Ajzen I. From Intentions to Actions: A Theory of Planned Behavior[M]. Berlin: Springer-Verlag, 1985.
[7] Krueger NF, Carsrud AL. Entrepreneurial intentions: Applying the theory of planned behaviour. Entrepreneurship & Regional Development, 1993, 5(4): 315-330. doi:10.1080/08985629300000020
[8] Krueger NF, Reilly MD, Carsrud AL. Competing models of entrepreneurial intentions. Journal of Business Venturing, 2000, 15(5-6): 411-432. doi:10.1016/S0883-9026(98)00033-0
[9] Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory[M]. Englewood Cliffs: Prentice-Hall, 1986.
[10] Shapero A, Sokol L. The social dimensions of entrepreneurship[M] //KENT C A, SEXTON D L, VESPER K H. Encyclopedia of Entrepreneurship. Englewood Cliffs: Prentice-Hall, 1982: 72-90. https://www.scirp.org/reference/referencespapers?referenceid=3840736
[11] Bronfenbrenner U. The Ecology of Human Development: Experiments by Nature and Design[M]. Cambridge, MA: Harvard University Press, 1979. doi:10.4159/9780674028845-toc
[12] Martin BC, Mcnally JJ, Kay MJ. Examining the formation of human capital in entrepreneurship: A meta-analysis of entrepreneurship education outcomes[J]. Journal of Business Venturing, 2013, 28(2): 211-224. doi::10.1016/j.jbusvent.2012.03.002
[13] Zhang ZQ, Huang ZR. The impact of household capital on the entrepreneurial intention of undergraduate students in vocational universities: the chain-mediated role of social support and critical thinking. Psychological Monthly, 2024, 19(14): 72-75+94. doi::10.19738/j.cnki.psy.2024.14.020.
[14] Yue L, Shengnan P, Lu Z and Salarzadeh JH. The longitudinal psychological pathway from family support to entrepreneurial intention: a mediation model of entrepreneurial self-efficacy among art and design students. Front. Psychol. 2026, 17:1745833. doi: 10.3389/fpsyg.2026.1745833
[15] Liang J, Zeng YZ, Cao SL. Research on the Construction of University-Family-Society Collaborative Education Mechanism in Applied Universities. Advances in Education, 2026, 16(7): 106-114. doi:10.12677/ae.2026.1671345.
[16] Ning DP, He TT, Deng JX, Zeng X. A Study of the Mechanism Through Which Innovation and Entrepreneurship Education Influences Entrepreneurial Behavior: A Large-Sample Empirical Investigation Using Entrepreneurial Ability and Entrepreneurial Willingness as Mediators. Journal of East China Normal University (Education Science Edition), 2023, 41(2):93-105. doi:: 10.16382/j.cnki.1000-5560.2023.02.008

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)

© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261121

Configurational Pathways in Digital Textbook Development: A TOE-Based Analysis of Business English Materials for Vocational Education

Jun LIU [1]
Department of Foreign Languages, Shanghai Publishing and Printing College, China

Abstract. Digital textbooks (DTs) constitute core resources for competency-based vocational education. This study applies the Technology-Organization-Environment (TOE) framework configurationally to examine the development of digital Business English materials, shifting the analytical lens from variable-centered approaches to configurational logic that treats technological, organizational, and environmental factors as interdependent conditions. Drawing on a written adaptation of autobiographical narrative inquiry with 11 stakeholders from a Chinese higher vocational institution, and using social network analysis as a structuring heuristic, the study identifies 27 configurational factors across nine TOE elements and analytically derives three equifinal pathways, technology-concentrated, organization-centered, and environment-linked, labeled by domain concentration rather than causal direction. Cross-domain coupling indices reveal robust technology-organization co-occurrence (density=0.34) and sparse organization-environment linkages (density=0.15); narrative evidence interprets this asymmetry as a governance bottleneck rooted in incentive asymmetry and institutional hollowness. The study further proposes a four-stage optimization framework and a testable proposition regarding stakeholder-perceived student engagement as a mediating feedback signal. Theoretically, it offers a project-level configurational application of TOE and introduces a narrative-SNA heuristic for small-N, context-rich configurational inquiry. Practically, it reframes the policy challenge from resource insufficiency to governance mechanisms, providing ecosystem-aware guidance for digital materials development in vocational Business English contexts.

Keywords. digital textbooks, TOE framework, configurational analysis, vocational Business English, cross-domain coupling

1. Introduction

The global shift to digital textbooks (DTs) represents a defining movement in educational digitization. However, its success hinges not on technology alone but on the interplay of technological, organizational, and environmental factors [1][2]. This is particularly salient in vocational education and training (VET), where curricula must keep pace with evolving workplace competencies. For Business English materials, a

[1] Corresponding author: Jun Liu, 5100 Puxing Road, Fengxian District, Shanghai 201404, China. E-mail: jun.liu@sppc.edu.cn

field integrating language proficiency, intercultural communication, and professional practice, DTs offer distinctive affordances by delivering authentic multimodal scenarios and supporting adaptive learning trajectories [3][4].

Despite technological progress, empirical evidence consistently shows that DT effectiveness depends on coordinated alignment among technical capacities, organizational resources, and contextual demands [5][6]. Nevertheless, prevailing research exhibits three persistent limitations. First, variable-centered studies examine these influences separately, overlooking the interdependence central to DT development [7]. Second, existing research prioritizes post-adoption evaluation over developmental processes, treating DTs as static products rather than evolving ecosystems [8]. Third, research on Business English materials remains limited, and generic frameworks inadequately capture the dual demands of language pedagogy and industry alignment [4][9].

To address these gaps, this study employs the Technology-Organization-Environment (TOE) framework [10] as a configurational lens. This shifts the analytical question from which factors matter to how combinations of factors interact to produce outcomes, aligning with configurational principles of equifinality and conjunctural causation [11]. Two sets of questions guide the inquiry:

Q1: What configurational factors across TOE domains shape the development of digital Business English materials for vocational education?

Q2: (a) What heuristic configurational pathways underpin high-quality development? (b) What cross-domain coupling patterns characterize the stakeholder ecosystem?

Three core terms are defined for conceptual clarity. Configurational factors refer to conditions across TOE domains that collectively shape outcomes. Pathway denotes a specific combination of such factors associated with high-quality development, derived via interpretive analysis. Coupling refers to the density of perceived cross-domain linkages measured through narrative co-occurrence. Consistent with the title, “Business English materials” in this study refers specifically to digital materials. This study distinguishes core anchor conditions, the baseline factors shared across all pathways, from pathway-specific focal conditions, the domain-salient factors that organize each configuration; these focal labels describe relative prominence, not causal direction

Using a written adpation of autobiographical narrative inquiry and social network analysis (SNA) at a Chinese VET institution, this study offers four contributions: (1) a project-level configurational application of TOE; (2) three equifinal labeled by domain concentration rather than causal direction; (3) a four-stage optimization framework; and (4) a narrative-grounded diagnosis of sparse O-E coupling as a governance bottleneck. The China’s VET policy context, including the Dual High Plan and quality initiatives that link industry-education integration to evaluation and funding, intensifies pressure for sustained cross-sector collaboration.

2. Literature Review

2.1 Digital Textbook Development: Theoretical Evolution and Global Practice

Scholarly inquiry into digital textbooks has evolved through foundational conceptualization, functional expansion, and ecological integration. Early work defined DTs as integrated systems encompassing interactive content, multimedia, assessment

tools, and adaptive functionalities, distinguishing them from digitized print materials [12]. Subsequent stage-based models described a progression from static format conversion to platform-based artificial intelligence (AI) systems with adaptive learning capabilities [13][14].

Contemporary research has shifted toward ecological perspectives, emphasizing that DT quality is shaped by multi-stakeholder governance, institutional support, and contextual fit [15][16]. Global bibliometric mapping indicates an evolutionary shift from foundational efficacy testing to AI-enabled resource design, with East Asian higher education research expanding rapidly under national digital education policies [17]. Although generative AI promises to restructure DT ecosystems, empirical implementation evidence remains limited [18][19].

Within higher education, a hybrid acceptance framework rooted in self-determination theory [20] indicates that student autonomy, competence, and relatedness strongly predict learning engagement and academic performance. Disciplinary disparities also persist: medical and science fields adopt DTs far more widely than the humanities, as disciplinary norms differentiate resource development paths. Adjacent educational-digitalization studies indicate that effective digital transformation requires staged teacher digital-literacy development and institutional support [21], while large-language-model teaching systems and AI-based classroom or motor-skill assessment increasingly convert learner-state and behavioral data into adaptive feedback [22][23][24][25].

Cross-national comparisons reveal divergent DT governance regimes constrained by economic levels, infrastructure, and policy intensity [26]. Three archetypal policy modes emerge: proactive, reactive, and ad hoc. High-income jurisdictions including Singapore and the United States deliver greater curricular equity via mandatory content reviews, inclusive procurement, and sustained teacher training; low- and middle-income economies tend to distribute hardware without mitigating cultural or connectivity barriers. Governance centralization and infrastructure investment account for 70% of cross-country outcome variance.

Chinese practice forms three segmented implementation pathways by educational stage: iterative updating for basic education, knowledge graph embedding for higher education, and job-curriculum alignment for vocational education [27]. However, VET deployment suffers fragmentation, marked by poor cross-stakeholder coordination and unsustainable resource iteration [8]. Unlike the decentralized incentive-audit system of the United States and Singapore's centralized real-time oversight mechanism, Chinese VET operates under a hybrid model balancing institutional autonomy and top-down national guidelines. This structure tends to prioritize technological rollout while weakening organization-environment coordination, whose long-term sustainability effects require deeper investigation.

2.2 Vocational Business English and the Need for Integrated Perspectives

Vocational Business English falls within the established frameworks of English for Specific Purposes, emphasizing the integration of language proficiency and domain-specific knowledge acquisition [3][4]. The vocational Business English classroom sits at the intersection of multiple knowledge systems. Learners develop linguistic proficiency, professional communication skills, cross-cultural pragmatic awareness, and workplace literacy such as commercial correspondence drafting and negotiation. This multidimensional mandate distinguishes it from general English

instruction, which focuses on everyday communicative competence, and from non-language vocational education, where linguistic demands are secondary to technical skills. In vocational Business English, language and professional practice are fused in authentic workplace scenarios requiring integrated communicative action.

This pedagogical integration imposes corresponding requirements for textbook development. A vocational Business English DT cannot function merely as a digitized language syllabus; it must simultaneously serve as a language learning resource, a professional competency scaffold, a workplace communication simulator, and an adaptive assessment tool [4]. Delivering these functions requires coordinated input from stakeholders with distinct expertise. Language educators contribute pedagogical content knowledge, publishing professionals provide technological infrastructure and quality assurance, and industry practitioners supply authentic workplace scenarios and competency standards. Successful integration therefore relies on cross-stakeholder communication, incentive alignment, and institutional mechanisms for sustained collaboration.

Nevertheless, the literature provides limited guidance on accomplishing this integration. Research on language textbook design, digital resource development, and industry-education collaboration remains largely parallel [8], yielding fragmented insights rather than a lifecycle account of how these components interact. This gap is particularly consequential for Business English materials, whose value depends on integrating language pedagogy, digital technology, and professional practice. Because these demands span technological, organizational, and environmental dimensions, TOE offers a suitable lens for theorizing their integration.

2.3 The TOE Framework and Configurational Analysis in Educational Research

The TOE framework [10] maintains that technology adoption outcomes emerge from the interplay of three contextual domains: technological (existing technologies and organizational capabilities), organizational (structural characteristics and resource endowments), and environmental (institutional pressures and stakeholder relationships). Its core proposition that adoption is shaped by cross-domain alignment rather than technical attributes alone makes it a flexible analytical tool for organizational and educational research [28]. Unlike reductionist models that isolate individual predictors, TOE inherently acknowledges dimensional interdependence, making it suitable for multi-stakeholder innovations requiring cross-domain coordination.

TOE applications have expanded widely in educational technology research, covering learning management platforms and digital instructional materials. Systematic reviews show that scholars disproportionately examine technical barriers, while organizational and environmental constraints remain underinvestigated despite their vital mediating roles in institutional implementation [28]. Most existing TOE studies rely on variable-centered methods including regression and structural equation modeling to assess isolated factor effects, an approach that overlooks the framework's systemic logic. If the three TOE domains are interdependent, their impacts should be analyzed via configurational condition combinations rather than additive linear estimation.

The configurational application of TOE in this study differs from variable-centered approaches in three ways. First, it maps equifinal condition combinations rather than estimating the largest net effect. Second, it treats factors as interdependent conditions whose influence depends on complements. Third, it diagnoses systemic imbalance

rather than average effects. This is a configurational application, not a reconceptualization: TOE's three domains remain intact, but conditions are analyzed as conjunctural and equifinal. This approach suits DT development, where shortfalls often reflect missing complementary inputs rather than insufficient aggregate resources.

3. Methodology

3.1 Research Design

This study employed a written adaptation of autobiographical narrative inquiry to examine the configurational dynamics of vocational Business English DT development. Narrative inquiry is suited to examining complex, context-embedded organizational phenomena through insiders' lived experiences [29][30]. Unlike oral interview-based inquiry, this adaptation does not claim full relational co-presence. Instead, it retains first-person, temporally ordered accounts while treating written texts as the primary field texts. The written mode enabled paced retrospective composition and comparable TOE coverage across stakeholder groups, while follow-up clarification provided a bounded relational check. First-person accounts reveal the interconnected, conjunctural character of outcome-shaping factors, aligning the approach with configurational thinking [31].

The study adopted an interpretive configurational stance, treating narratives as sense-making representations that organize experience into condition-outcome configurations. SNA served a strictly heuristic function. Although density, centrality, and modularity originated in a positivist tradition of structural discovery, narrative inquiry prioritizes meaning interpretation. In this study, SNA outputs organized and visualized co-occurrence patterns to guide interpretation rather than test hypotheses or estimate population parameters [31]. Each metric was treated as a descriptive guide to potentially meaningful configurations and re-anchored in narrative semantics; substantive conclusions drew on validated narrative content rather than metric magnitude alone.

Two epistemological boundaries were observed. First, SNA was used not for causal inference but to identify co-occurrence structures among narrated factors; its function was to generate analytical insights about configurational pathways rather than to test causal relationships. Second, the subnetworks produced by community detection were treated as analytical constructs requiring semantic validation through qualitative content analysis, rather than as ontological entities or formally verified sufficient condition sets. Their validity rests on interpretive consensus established through member checking and peer debriefing. Accordingly, this study uses calibrated verbs such as *identified*, *suggested*, and *analytically derived* when describing results, reserving causal language for participants' own narrated accounts.

The research was conducted at a public higher vocational institution in eastern China (designated Institution S), which has implemented systematic DT development for Business English practical training since 2022. The initiative comprises four development teams, each consisting of Business English instructors, professional publishing editors, and industry experts from international trade enterprises. Teams collaborate to produce DTs aligned with national VET competency standards and workplace requirements. Ethical approval was obtained, and all participants provided written informed consent.

3.2 Participants

Participants were recruited via purposive sampling to ensure representation of all stakeholder groups involved in DT development. The final sample included 11 participants across three role categories: seven textbook authors (Group A, coded A1-A7), two publishing editors (Group B, coded B1-B2), and two industry experts (Group C, coded C1-C2). All participants had direct, ongoing involvement in the DT project and a minimum of three years of relevant professional experience.

This sample size reflects a deliberate trade-off between analytical depth and role coverage. Larger samples would sacrifice the contextual richness required to capture stakeholders' lived experiences in a specific DT project. The study prioritized an information-rich sample to enable sustained engagement with each participant's account, facilitating the thick description essential to configurational narrative inquiry.

The role distribution corresponds directly to the composition of the DT development teams. The two industry experts constitute the population of external enterprise partners, ensuring the environmental-domain perspective is exhaustively captured. This tripartite composition was theoretically motivated, since successful DT development in vocational Business English requires coordinated input from language pedagogy, publishing infrastructure, and professional practice standards.

3.3 Data Collection

Data were collected via written semi-structured autobiographical narrative protocols organized around the TOE domains and two developmental phase pairs: compilation/publishing and review/adoption. Protocol development proceeded in three stages: drafting from TOE and DT literature; piloting with two Business English instructors outside the final sample; and review by an independent educational research methodologist. Universal equivalence between written and oral narratives was not assumed. Comparability was purpose-bound: prompts elicited the event sequences, perceived dependencies, and cross-domain linkages required for configurational coding, while follow-up clarification and member checking resolved ambiguous chronology or interpretation. Because written responses provide fewer prosodic cues and less spontaneous relational co-construction, coding was restricted to explicitly stated content.

The final protocol included 12 prompt sets, each consisting of a primary open-ended question, one to two probing questions, and a reflective prompt. Prompt design was guided by the need to elicit configurational narrative data, accounts revealing not just isolated factors but their interconnections. Prompts explicitly invited participants to describe sequences of events, dependencies among resources and outcomes, and cross-stakeholder interactions. Example prompts included: describing initial involvement and how expectations evolved through stakeholder interactions; reflecting on technological resources and challenges that shaped contributions, and identifying organizational or external conditions that enabled or constrained access; and discussing how organizational support structures and external stakeholder relationships influenced the development process, with specific instances of cross-domain factor interaction.

Protocols were distributed in staggered batches. Participants submitted written narrative responses within a three-day window; follow-up clarification conversations were used only to resolve ambiguity and verify event chronology, not to convert the

texts into oral interviews. Ten complete narrative texts were collected. One industry expert (C2) provided partial responses, completing the compilation-phase prompt sets but not the review-and-adoption phase sets. Consequently, only explicitly stated factors from C2's completed segments were admitted into the qualitative thematic analysis, where they informed the interpretation of environment-domain factors; C2 was excluded from all frequency-count denominators and from weighted network edge calculations to avoid estimation bias. The content boundaries of C2's responses are documented in the audit trail.

3.4 Data Analysis

Data analysis proceeded through three integrated phases aligned with the configurational theoretical framework.

Phase 1: Theoretically Informed Coding. Narrative texts were coded using a three-level hierarchical structure grounded in the TOE framework: core TOE domains, interlocking elements within each domain, and granular configurational factors. The coding structure was developed deductively but allowed inductive refinement. Factor granularity followed explicit rules: a candidate factor was split only when at least two participants articulated functionally distinct sub-dimensions with distinguishable consequences; statements were merged when semantically interchangeable across participants. Saturation was assessed at the factor level: no new factors appeared after the ninth participant, and no new interlocking elements after the seventh, indicating saturation of the 27-factor set. Two researchers independently coded all narratives, achieving substantial inter-coder agreement (Cohen's κ=0.84; per-factor range: 0.79-1.00). Discrepancies were resolved through collaborative discussion.

Phase 2: Qualitative Content Analysis. Coded data were synthesized to identify recurring patterns, stakeholder consensus, and areas of tension across TOE domains. This phase involved iterative reading, thematic clustering, and interpretive analysis of factor interactions, serving as the interpretive bridge between raw coding and network construction. Co-occurrence relationships were identified and validated during this phase, ensuring subsequent SNA was grounded in substantive interpretive judgment.

Phase 3: Social Network Analysis. Gephi (version 0.10.1) was used to construct and analyze a configurational factor network with 27 nodes. Edges connected factors under two rules: narrative proximity (co-occurrence within the same paragraph) and explicit causal or dependency language, identified via a predefined lexical protocol. Purely additive listings were excluded. Both coders independently applied these rules; initial agreement on candidate edges reached 91.3%, with disputed cases resolved through joint re-reading or conservative exclusion.

Network construction followed a four-step protocol. Nodes retained full qualitative definitions to prevent semantic drift. Edges were weighted by the number of participants supporting each linkage (integer weights 2-10) rather than treated as binary, mitigating over-amplification of weak ties in small samples. A minimum frequency threshold of 2 eliminated single-source idiosyncrasies, removing 12 low-frequency candidate edges (15.4%). Systematic sensitivity analyses confirmed stability: adjusting the proximity unit produced stable top-five degree centrality rankings (Spearman's ρ>0.92); granularity merging left community structure unchanged; bootstrapped edge-resampling reproduced the Louvain community structure in 94.7% of iterations; and leave-one-out node-deletion recovered the same three-community solution in every iteration.

Degree centrality (hub status) and betweenness centrality (bridging roles) were calculated. Network density equaled observed edges divided by maximum possible edges. For cross-domain coupling, denominators were the products of domain node counts (10 Technology, 9 Organization, 8 Environment nodes), yielding maxima of 90, 80, and 72 possible T-O, T-E, and O-E edges. Coupling densities were computed on the thresholded dichotomized network to represent connection breadth independent of tie strength; weighted variants preserved the ordering. Louvain community detection (default resolution) identified three communities (Q=0.41) Centrality and community results derive from the weighted network, whereas coupling densities use the the dichotomized backbone.

3.5 Reliability and Validity

Reliability and validity are addressed by adopting Lincoln and Guba's criteria [32] in qualitative research. Credibility was enhanced through 18-month engagement with the research context, triangulation across three stakeholder groups, member checking with six participants, including verification of event chronology and cross-factor linkages in the written accounts, and regular peer debriefing. Triangulation across author, editor, and industry expert accounts allowed cross-verification of reported events and identification of divergent perspectives, treated as analytically informative rather than as errors.

Transferability was supported through thick description of the institutional context, participant characteristics, and analytical procedures. Dependability was maintained via a detailed audit trail documenting all coding decisions and analytical revisions. Confirmability was strengthened through reflexive journaling, in which researchers explicitly documented and bracketed prior assumptions. Regular peer debriefing with colleagues not involved in data collection served as an additional check against interpretive bias.

4. Results

4.1 Q1: Configurational Factors Across TOE Domains

Table 1 maps all 27 configurational factors across nine interlocking TOE elements, with total-mention frequencies indicating overall salience. Within the Technology domain, digital literacy emerged as the most salient element. DT development capability (T3-a) received unanimous endorsement (frequency=10), with narratives describing project participation as transformative for building multimedia authoring, interactive content design, and assessment integration competencies. The international business case library (T2-b, frequency=9) was identified as a defining feature of high-quality Business English materials, with participants emphasizing that authentic workplace scenarios differentiated vocational materials from general language resources. Infrastructure factors received comparatively less attention, indicating they were viewed as necessary but insufficient.

Within the Organization domain, resource input factors emerged as critical enablers. Publishing project funds (O2-b, frequency=7) and training investment (O2-c, frequency=6) were consistently linked to development quality. Editors emphasized digital publishing's substantially higher resource intensity. Authors described targeted

training as transformative. Evaluation systems represented a notable gap: participants consistently reported the absence of standardized DT quality criteria (O3-a) and insufficient ongoing effectiveness feedback mechanisms (O3-b).

Within the Environment domain, teaching community factors emerged as most prominent. Joint development seminars (E3-a) received universal endorsement (frequency=10), described as generative spaces for knowledge sharing, content refinement, and cross-stakeholder alignment. Industry practice competency demands (E1-b, frequency=8) were consistently highlighted, with industry experts detailing growing gaps between traditional textbook content and real-world workplace requirements. However, participants noted sustainability challenges in enterprise engagement, with industry participation often limited to initial project phases.

Table 1. TOE Configurational Factors in Digital Textbook Development

Domain	Interlocking Element	Representative Factors (Participant Sources)	Unique Participant Mentions (min–max per sub-factor)
Technology (T)	Infrastructure	T1-a: Platform construction (A1, A4, A6, B1); T1-b: Technical development (A2, A3, A7, B2, C1)	4-5
	Digital Resource Supply	T2-a: Specialized resource development (A1, A3, A4, A5, B1); T2-b: International business case library (A1, A2, A3, A4, A5, A6, A7, B1); T2-c: Resource updating (A2, B1, B2)	6-9
	Digital Literacy	T3-a: Development capability (A1, A2, A3, A4, A5, A6, A7, B1, B2, C1); T3-b: Teaching capability (A1, A3, A7); T3-c: Evaluation capability (A2, A4, A5); T3-d: Platform editing (A2, A3, A6); T3-e: Learner digital skills (A1, A2, A3, A4, A5, A6, B1, B2, C1)	5-10
Organization (O)	Attention Allocation	O1-a: Professional development mechanisms (A1, A2, A3); O1-b: Research support (A1); O1-c: Learner training (A5, B1, B2); O1-d: Copyright management (A5)	1-4
	Resource Input	O2-a: Technology investment (A3, A4, A6); O2-b: Publishing funds (A1, A3, A4, A5, B1, B2, C1); O2-c: Training investment (A1, A2, A3, A7, B1)	3-7
	Evaluation	O3-a: Standards and systems (A1, A5, A7, B2); O3-b: Effectiveness feedback (A2, A3, A5, A6, A7, B1, B2)	5-7
Environment (E)	Learning Needs	E1-a: New ecosystem requirements (A1, A4, A5, A7, B2); E1-b: Industry practice demands (A1, A2, A4, A5, A6, A7, C1)	6-8
	Enterprise Resources	E2-a: Application scenarios and data (A2, A6); E2-b: Expert participation (A1, A3, A4, A5, B1)	2-5
	Teaching Community	E3-a: Joint development seminars (A1, A2, A3, A4, A5, A6, A7, B1, C1); E3-b: Enterprise feedback (A7, B2, C1); E3-c: Publisher feedback (A1); E3-d: Community management (A4, A5, C1)	2-10

Note. Values in the final column indicate the minimum and maximum number of unique participants mentioning any sub-factor within each interlocking element; a value of 10 denotes mention by all participants who provided complete narrative responses (C2 submitted partial responses and was excluded from frequency counts).

4.2 Q2(a): Three Configurational Pathways of Development

Community detection (Louvain, Q=0.41) organized coded factors into three candidate subnetworks (Figure 1), treated as analytical constructs whose meaning and labels were semantically validated. Because the direction-free co-occurrence network provided no independent causal-attribution counts, pathways were labeled by relative domain concentration. The domain-salient factor in each community was designated the pathway-specific focal condition.

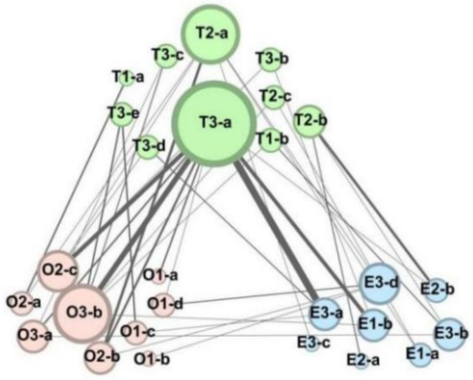


Figure 1. *Social Network Analysis of Configurational Factors*

Note. Node size corresponds to weighted degree centrality (larger nodes indicate more connections to other factors).

Pathway T: Technology-Concentrated. Predominantly associated with T3-a (development capability), the highest-degree node (degree=12), functioning as a central hub connecting all three TOE domains. Additional core nodes included T1-a (platform construction), T2-a (specialized resource development), and T3-e (learner digital skills). This pathway grouped accounts in which technological capability was the focal concern; the label does not imply that technology caused the outcome. T3-a required conjunctural support from O2-b and E3-a to translate technical skill into high-quality outcomes. This pathway exhibited the strongest internal cohesion (density=0.38).

Pathway O: Organization-Centered. Characterized by O2-b (publishing funds, degree=8) and O2-c (training investment, degree=7), functioning as bridging nodes connecting organizational resource provision to technological and environmental domains. This pathway described contexts where resource availability defined development scope and quality. The high betweenness values locate organizational factors at frequent narrative junctions; they do not establish that those factors mediated outcomes.

Pathway E: Environment-Linked. E3-a (teaching community engagement, degree=9) and E1-b (industry competency demands, degree=8) were the most prominent environment-domain nodes. This pathway grouped accounts emphasizing external stakeholder connections; it does not claim that external pressure directed the other domains. E-pathway nodes formed a cohesive cluster (density=0.32) but exhibited the lowest betweenness centrality overall, suggesting environmental factors were more locally influential than globally bridging.

The three pathways shared three core anchors: T3-a, O2-b, and E3-a appeared across all configurations. Pathway distinctiveness arose from which domain-dominant

factor served as the pathway-specific focal condition and how supporting conditions were configured around the anchors. Table 2 summarizes comparative profiles.

Table 2. Comparative Configurational Profiles of Three Pathways

Pathway	Pathway-Specific Focal Condition	Shared Core Anchors	Internal Density	Illustrative Narrative Sources
T: Technology-Concentrated	T3-a (Development Capability)	O2-b, E3-a	0.38	A2, A3, A6
O: Organization-Centered	O2-b (Publishing Funds)	T3-a, E3-a	0.31	A1, A4, A5
E: Environment-Linked	E3-a (Joint Seminars)	T3-a, O2-b	0.32	C1, B2

Pathway-Outcome Linkage and Narrative Validation. Because independent quality ratings were unavailable, pathway-outcome linkage was assessed through narrative semantics. Each pathway contained at least three nodes associated with favorable evaluations. Condensed trajectories for three illustrative case sets (Table 3) preserved narrators' temporal-causal ordering and coherently mapped onto the algorithm-derived pathways. They corroborate pathway coherence but do not establish causal direction. Cases were selected purposively: each mentioned at least two relevant community nodes, included the universal anchor T3-a, and contained an explicit outcome statement. Unselected narratives also shared T3-a and other anchors but varied in community-node coverage; the trajectories are therefore illustrative, not exhaustive.

Table 3. Illustrative Composite Narrative Trajectories Mapped to the Three Configurational Pathways

Pathway	Condensed Narrative Trajectory	Sources	Outcome as Narrated
T: Technology-concentrated	Limited authoring skills→ embedded platform training→ first interactive unit produced→ successful bid for additional funds→ expanded industry case partnerships	A2, A3, A6	Units passed expert review; team secured further project funding
O: Organization-centered	Earmarked publishing grant→ multimedia features commissioned→ targeted training expanded team capability→ broader case and assessment repertoire	A1, A4, A5	Textbook adopted across multiple programs; sustained revision cycle
E: Environment-linked	Enterprise critique of outdated cases→ joint seminars redirected priorities→ platform and content redesign→ competency-aligned assessment added	C1, B2	Employer satisfaction; renewed enterprise commitment

Note. Composite trajectories are condensed from the autobiographical narratives and preserve the narrated sequence expressed by the narrators; they are presented as interpretive validation of the algorithm-derived communities, not as independent causal evidence.

4.3 Q2(b): Cross-Domain Coupling Patterns

Descriptive co-occurrence indices summarized distinct coupling patterns across the three TOE domain pairs, with marked variation in connection density; these indices describe the breadth of coded narrative connections, not causal strength. Coupling density equals observed cross-domain edges divided by maximum possible edges.

T-O coupling was the densest pattern (31 observed of 90 possible edges; density=0.34), with frequent reciprocal co-occurrences between digital literacy factors and organizational resource inputs. The most frequent cross-domain co-occurrences were between development capability (T3-a) and training investment (O2-c), and between specialized resource development (T2-a) and publishing funds (O2-b).

Narratives portrayed investments in training, technology, and project funding as mutually reinforcing support for technological capabilities.

T-E coupling was moderately dense (17 of 80 possible edges; density=0.21), with linkages concentrated between digital resource development factors and industry resource and competency factors. In participant narratives, joint development seminars (E3-a) served as the principal channel through which industry knowledge and competency demands entered technological content and platform development decisions. This pattern suggests that environmental input was often channeled through organizational structures and collaborative community platforms.

O-E coupling was the least dense pattern (11 of 72 possible edges; density=0.15), with limited connections concentrated between attention allocation factors and teaching community governance. The most notable O-E linkages were between copyright management (O1-d) and community governance (E3-d), and between publishing funds (O2-b) and long-term enterprise engagement. The sparse O-E co-occurrence pattern, interpreted alongside participants' accounts, points to a critical developmental gap: organizational processes and incentive systems were not effectively aligned with environmental stakeholder demands, particularly for sustaining long-term industry collaboration. This gap limited the ecosystem's capacity to respond dynamically to evolving industry competency requirements.

5. Discussion

5.1 Project-Level Configurational Application: Pathways and Core-Periphery Structures

The three pathways, namely technology-concentrated, organization-centered, and environment-linked, provide empirical grounding for a configurational application of the TOE framework at the development-project level. Prior variable-centered TOE research has evaluated the predictive impact of separate factors independently [28], implicitly treating domains as additive contributors rather than interdependent conditions. The present findings indicate that developmental quality stems from combined contextual conditions rather than standalone drivers, aligning with recent configurational scholarship in education [6][33]. This alignment extends the TOE framework's scope from technology adoption to developmental process research, an application that has received limited attention in the digital textbook literature.

The three pathways shared three anchors: T3-a (development capability), O2-b (publishing funds), and E3-a (joint seminars) appeared across all configurations Pathway distinctiveness arises from which domain-dominant condition serves as the pathway-specific focal condition and how supporting conditions are configured around the anchors. This core-periphery pattern is consistent with configurational theory's distinction between core and peripheral conditions [7].

This application differs from existing configurational TOE studies in three respects: the unit shifts from national systems or institutions[6][34] and individual innovation capability[38] to the development project; the method shifts from fuzzy-set qualitative comparative analysis (fsQCA)[6][28][35] to a narrative-SNA heuristic suited to small-N, context-rich settings; and the output shifts from sufficient-condition sets to coupling-density diagnosis for locating governance intervention.

The pathway-endowment correspondence suggests that fit depends on baseline TOE conditions. Digital maturity may favor technology-concentrated pathways, resource centralization organization-centered pathways, and mature partnerships environment-linked pathways. This pattern is consistent with configurational accounts of path dependence: institutions should design development strategies that build on existing strengths rather than mandating uniform restructuring.

5.2 Cross-Domain Coupling Asymmetry: The O-E Bottleneck as a Governance Challenge

The descriptive coupling pattern revealed a structural asymmetry with significant theoretical and practical implications. Narratives associated with dense T-O co-occurrence indicate well-internalized institutional synergies. Organizational investments in training, technology, and project funding effectively target technological capability building, creating a reinforcing cycle. This pattern is consistent with prior TOE-based research in Chinese vocational education, where organizational resource support has been identified as the strongest predictor of technological implementation success [6][34][35].

By contrast, the sparse O-E co-occurrence pattern (density=0.15), read together with narratives of weak alignment, represents a critical systemic gap. Organizational processes and incentive systems are not effectively aligned with environmental stakeholder demands, particularly for sustaining long-term industry collaboration. This finding echoes broader research on Chinese VET industry-education collaboration, which has documented persistent challenges in establishing mutually beneficial partnerships [8]. Accordingly, SNA maps where narrative connections are structurally thin; the substantive diagnosis of governance failure derives from participants' accounts and prior collaboration research, not from the density value alone.

Two narrated sub-mechanisms help explain this deficiency. First, incentive asymmetry: institutions and society capture most returns to enterprise participation, whereas enterprises bear the costs. Without shared intellectual property or prioritized talent pipelines, continued participation may be organizationally irrational. Second, institutional hollowness: policy mandates school-enterprise cooperation but rarely specifies engagement duration, contribution measurement, or deliverable-sharing rules. Collaboration therefore depends on personal relationships rather than durable commitments, helping explain persistent fragmentation despite policy initiatives.

Within the global policy framework [26], the Chinese VET DT governance model can be classified as a proactive-reactive hybrid. National top-down policy mandates drive proactive investment in technological infrastructure and organizational funding, producing strong T-O coupling. By contrast, institutional-level industry collaboration mechanisms remain reactive and ad hoc, directly accounting for weak O-E coupling. This reading bridges national policy typology research with meso-level implementation analysis.

5.3 Stakeholder-Perceived Student Engagement: Feedback Signal and Theoretical Proposition

Two distinct roles of student-related perceptions must be separated conceptually. As a development input, stakeholders' perceptions of student needs and digital skills (T3-e, degree=10) shaped design decisions. As a development output, stakeholders'

perceptions of student engagement functioned as a signal through which teams judged whether configurations were working. The data capture the former directly and the latter only through adult stakeholders' perceptions, reflecting the marginalization of student perspectives identified in prior research [36].

Perceived student engagement is closely tied to perceived workplace relevance in vocational Business English. When textbook content aligns with industry competency requirements, learners are more likely to perceive instructional activities as meaningful for professional development. The configurational analysis indicates that this alignment is associated with the interaction of T2-b (authentic case development), E1-b (industry competency demands), and E3-b (multi-stakeholder feedback). This configuration spans all three TOE domains and cannot be produced through interventions confined to any single domain.

The reasoning is formalized as a testable proposition: Stakeholder-perceived student engagement functions as a mediating feedback signal between TOE configurations and the long-term sustainability of DT development. Configurations characterized by stronger cross-domain coupling are expected to be associated with higher perceived engagement. Higher perceived engagement, in turn, is hypothesized to reinforce stakeholder resource commitment and iterative refinement, thereby contributing to developmental sustainability [6][35]. This proposition positions engagement as a feedback mechanism rather than a static outcome, consistent with calls to move beyond post-hoc evaluation toward developmental process analysis [6][35].

Adjacent studies clarify the value and limits of this signal. Teacher digital-literacy research shows that sustainable technology use requires phased training, technical support, and institutional incentives rather than access alone [21], explaining why T3-a recurred as an anchor. Adaptive digital-human systems show how learner-state data can personalize feedback [22], while AI-based classroom and movement assessment converts behavioral evidence into actionable evaluation [23][24][25]. However, these systems remain constrained by data quality, contextual adaptation, and deployment resources [22][23][24][25]. Intelligent feedback therefore strengthens DT development only when technological capability, organizational support, and stakeholder governance are jointly configured.

The proposition also has implications for the O-E coupling diagnosis. If perceived engagement indeed mediates the translation of TOE configurations into sustained commitment, then weak O-E coupling may partially explain why industry partners remain disengaged. Without visible signals of learner engagement flowing back to enterprises, the returns to participation remain unobservable, reinforcing incentive asymmetry. This reading suggests that improving feedback transparency could serve as a practical mechanism for strengthening O-E coupling, aligning with policy emphasis on integrating student voice throughout digital resource development [37]. The proposition awaits direct validation with student-reported data, and its testing constitutes a priority for future research.

6. Implications

6.1 Stage-Specific Development Strategies

During compilation, institutions should embed progressive digital literacy training within ongoing compilation work, paired with communities of practice to strengthen

T3-a. Long-term enterprise partnership frameworks should be formalized to sustain access to authentic workplace scenarios (PR). Joint development seminars should be institutionalized as a mandatory core process, creating a stable boundary-spanning venue for knowledge exchange (E3-a).

During review, static inspection should shift to multidimensional quality assurance covering technical functionality, pedagogical alignment, ethical compliance, and industry competency relevance. Review panels should include instructors, publishers, industry experts, and students. Dynamic feedback should integrate user experience and learning outcome data into continuous iteration, addressing the O3-b gap.

During publishing, technical infrastructure and governance should address the O-E bottleneck. Funding should prioritize demonstrated cross-domain TOE alignment. Investment should expand intelligent publishing infrastructure, including automated formatting, dynamic updating platforms, and copyright management. Expert advisory boards and shared IP frameworks should mitigate incentive asymmetry.

During adoption, institutions should align DTs with local contexts. A TOE diagnostic tool should assess the suitability across the three domains. Perceived student engagement should be tracked as a leading process indicator through formal feedback channels to development teams. A cross-institutional best-practice repository should support pathway migration across endowment profiles.

6.2 Systemic Governance and Policy Coordination

At the policy level, the findings redirect priorities from standalone technological investment toward cross-domain coordination capacity building. The weak O-E coupling diagnosis indicates that interventions confined to technological or organizational domains are unlikely to resolve sustainability challenges. Systemic governance reforms addressing the structural disconnect between institutional processes and environmental stakeholder demands are required.

Policy frameworks should introduce formal value-sharing mechanisms for industry partners, including shared intellectual property rights, co-branded competency credentials, and prioritized talent recruitment pipelines. These mechanisms directly target the incentive asymmetry and institutional hollowness underlying the O-E coupling gap (TI). Institutional leaders should align performance assessment metrics with cross-domain coupling quality rather than narrow technological output indicators (ND). Evaluation systems that reward inter-domain coordination, sustained industry engagement, and demonstrated responsiveness to evolving competency demands would create organizational incentives for behaviors identified as critical for high-quality development.

Finally, configurational thinking should be integrated into professional development for all stakeholder groups (TI). Training programs should build shared awareness of cross-domain interdependence, enabling authors, editors, and industry practitioners to recognize how their individual contributions interact with factors in other domains to shape collective outcomes. This awareness supports the transition from discrete product development to ecosystem-oriented DT governance.

6.3 Implementation and Evaluation Considerations

The translation of these recommendations requires attention to implementation feasibility and contextual variation. Interventions should be tailored to existing

institutional strengths rather than applied uniformly. Institutions with established digital infrastructure may prioritize the technology-concentrated pathway; those with strong administrative coordination may pursue the organization-centered pathway; and those with mature industry partnerships may leverage the environment-linked pathway.

Implementation sequencing also merits consideration. The four-stage framework suggests a logical progression, but in practice, stages overlap and feedback loops operate across phases. Institutions should adopt flexible approaches allowing iterative adjustment based on emerging evidence.

Evaluation should track both process indicators and outcome measures. Process indicators include stakeholder participation rates in joint seminars, frequency of cross-domain communication, and perceived quality of collaboration. Outcome measures include stakeholder-perceived student engagement, content quality assessments, and sustained iteration cycles. These indicators collectively capture the configurational logic underlying high-quality development: process indicators reflect the presence of cross-domain interactions, while outcome measures reflect their effectiveness. Regular evaluation against these indicators would enable proactive identification of emerging coupling gaps and strategy adjustment.

7. Conclusion and Limitations

This study investigated the configurational dynamics of digital Business English materials development for vocational education through a configurational application of the TOE framework, combining a written adaptation of autobiographical narrative inquiry with social network analysis. The analysis identified configurational factors across the TOE domains and derived three equifinal pathways: technology-concentrated, organization-centered, and environment-linked. Descriptive co-occurrence indices summarized denser T-O coupling (density=0.34), moderate T-E coupling (0.21), and sparse O-E coupling (0.15); narrative evidence supports interpreting the sparse O-E pattern as a governance bottleneck reflecting misalignment between institutional processes and stakeholder demands. A four-stage optimization framework and a testable proposition regarding stakeholder-perceived student engagement as a mediating feedback signal were also proposed.

Three theoretical contributions are articulated. First, the study demonstrates a configurational application of the TOE framework at the development-project level, generating analytical constructs including cross-domain coupling density, pathway-specific configurations, and core-periphery condition structures that remain inaccessible under variable-centered specifications. Second, it introduces a written adaptation of autobiographical narrative inquiry combined with social network analysis as a configurational heuristic for small-N, context-rich research, complementing formal set-theoretic methods when such analysis is not feasible. Third, by integrating descriptive co-occurrence mapping with narrative interpretation, it identifies sparse O-E coupling as a structural bottleneck not reducible to resource constraints, shifting policy attention toward governance mechanisms, including shared intellectual property frameworks and sustained partnership incentives.

Several limitations delimit these contributions. The single-institution design restricts transferability, and the pathways require cross-contextual comparison. The interpretive stance treats SNA as a structuring heuristic, making the typology hypothesis-generating rather than hypothesis-confirming. The illustrative case sets do

not exhaust all community nodes and support semantic validation rather than statistical representativeness. The industry-expert subsample is small (N=2, one partial), potentially privileging institutional perspectives. The retrospective, cross-sectional design limits dynamic claims to the proposed framework. Written narratives reduced relational presence and paralinguistic cues; clarification and member checking only partially offset this limitation. The student-engagement evidence derives from adult stakeholder rather than student self-report. Finally, only relatively successful projects were analyzed, leaving low-quality configurations theoretically extrapolated.

Future research should pursue multi-institutional comparisons, longitudinal tracking, direct student-voice integration, enlarged industry samples, and multi-method validation including fsQCA. Despite these limitations, the study offers a theoretically grounded and methodologically innovative framework for understanding and guiding digital materials development in vocational Business English contexts.

References

[1] Rodríguez R, Bruillard E, Horsley M, Eds. Digital textbooks: What's new? Santiago de Compostela: Servizo de Publicacións da USC/IARTEM; 2015. 348 p.
[2] Organisation for Economic Co-operation and Development. OECD Digital Education Outlook 2023: Towards an Effective Digital Education Ecosystem [Internet]. Paris: OECD Publishing; 2023 [cited 2026 Jun 15]. 414 p. Available from: https://doi.org/10.1787/c74f03de-en.
[3] Blake RJ, Guillén G. Brave new digital classroom: Technology and foreign language learning. 3rd ed. Washington: Georgetown University Press; 2019. 248 p.
[4] Yang L. New-form college English textbooks: Connotation, development principles, and core features. Foreign Language World. 2024 Jan; (1):57-64.
[5] Reich J. Failure to disrupt: Why technology alone cannot transform education. Cambridge: Harvard University Press; 2020. 272 p.
[6] Shi W, Liu M. Analysis on configuration effect of driving factors of digital transformation of vocational education under the TOE theoretical framework. Contemporary Vocational Education. 2024 Dec; (6):20-27.
[7] Fiss PC. Building better causal theories: A fuzzy set approach to typologies in organization research. Academy of Management Journal. 2011; 54(2):393-420. doi: 10.5465/amj.2011.60263120.
[8] Xu Y. Vocational education digital textbook construction and application analysis. China Vocational and Technical Education. 2024 Jul; (7):16-26.
[9] Jiang L, Al-Shaibani GKS. Influencing factors of students' small private online course-based learning adaptability in a higher vocational college in China. Interactive Learning Environments. 2024; 32(3):972–993. doi: 10.1080/10494820.2022.2105901.
[10] Tornatzky LG, Fleischer M. The processes of technological innovation. Lexington: D. C. Heath & Company; 1990. 298 p.
[11] Misangyi VF, Greckhamer T, Furnari S, et al. Embracing causal complexity: The emergence of a neo-configurational perspective. Journal of Management. 2016; 43(1):255-282. doi: 10.1177/0149206316679252.
[12] Frydenberg J, Matkin G. Open textbooks: Why? What? How? When? Menlo Park: William and Flora Hewlett Foundation; 2007. 34 p.
[13] Holmes W, Bialik M, Fadel C. Artificial intelligence in education: Promises and implications for teaching and learning. Boston: Center for Curriculum Redesign; 2019. 242 p.
[14] Duarte da Costa BJ, Escarpini R, Silva L, et al. From digitized pages to adaptive platforms: A taxonomy of digital textbooks as enablers of technology-supported learning. In: ICERI2025 Proceedings; 2025 Nov; Seville, Spain. p. 3070–3078. doi:10.21125/iceri.2025.0982.
[15] Dobler E. E-textbooks: A personalized learning experience or a digital distraction? Journal of Adolescent & Adult Literacy. 2015; 58(6):478-487.
[16] Põldoja H. The structure and components for the open education ecosystem-Constructive design research of online learning tools [dissertation]. Aaltodoc: Aalto University. 2016; 130 p. https://urn.fi/URN:ISBN:978-952-60-6993-7
[17] Tang X, Chen S, Liu X. A bibliometric analysis of digital textbook research: Identifying trends, themes, and networks. IEEE Access. 2026;14:34998–35008. doi:10.1109/ACCESS.2026.3670174.

[18] Mishra P, Warr M, Islam R. TPACK in the age of ChatGPT and Generative AI. Journal of Digital Learning in Teacher Education. 2023; 39(4): 235-251. doi: 10.1080/21532974.2023.2247480.
[19] Petrova MG, Xu B. Generative AI and the evolution of interactive digital textbooks. In: 2025 International Conference on Educational Technology Management; 2025 Nov 7-9; Guangzhou, China. IEEE; 12-18. doi:10.1109/ICETM67477.2025.11413553.
[20] Ma T, Al-Adwan AS, Li N, Purwanto E, Meng W, Liang HN. Effective use of digital textbooks: A hybrid framework of acceptance and self-determination in higher education programming courses. IEEE Transactions on Education. 2025;68(3):312–321. doi:10.1109/TE.2025.3569558.
[21] Lan S. Research on the promotion strategy of teachers' digital literacy under the background of education digital transformation. International Scientific Technical and Economic Research. 2025; 3(2):136-147. doi:10.71451/ISTAER2531.
[22] Yang W. Real Time Adaptive Teaching Digital Human System Based on Large Language Model. International Scientific Technical and Economic Research. 2026; 4(2):166-185. doi:10.71451/ISTAER2620.
[23] Yao S. Application of YOLOv10 Integrated with Attention Mechanism in the Senseless Monitoring of Students' Classroom Psychological State. International Scientific Technical and Economic Research. 2026; 4(2):186-205. doi:10.71451/ISTAER2621.
[24] Li H, Sun Y. Research on Automatic Evaluation Algorithm of Students' Sports Action Standardization Based on Computer Vision. International Scientific Technical and Economic Research. 2026; 4(1):188-199. doi:10.71451/ISTAER2609.
[25] Lei H, Zhou M. Research on Automatic Evaluation Algorithm for Student Sports Movement Quality Based on Deep Learning. Journal of Discovery Core. 2026; 1(1):19-54. doi:10.67541/jdc2602.
[26] Rajendran M, Sabitkhanovich AA, Chokkalingam VK, et al. A comparative global study of board of education policy responses to digital textbook implementation and curricular equity. International Online Journal of Education and Teaching. 2025; 12(4):236-243.
[27] Li G, Wang Y, Wang J. Digital textbook institutionalization: Possibilities and approaches based on Japanese and Korean policy experiences. China Educational Information. 2024; 30(5):90-100.
[28] Isiaku L, Adalier A. Analyzing the barriers to blockchain adoption in educational sectors: A comprehensive literature review using the Technology-Organizational-Environment (TOE) framework. On the Horizon. 2025; 33(1):32-60. doi: 10.1108/OTH-05-2024-0029.
[29] Clandinin DJ, Connelly FM. Narrative inquiry: Experience and story in qualitative research. San Francisco: Jossey-Bass Publishers; 2004. 211 p.
[30] Pavlenko A. Autobiographic narratives as data in applied linguistics. Applied Linguistics. 2007; 28(2):163-188. doi: 10.1093/applin/amm008.
[31] Furnari S, Crilly D, Misangyi VF, et al. Capturing causal complexity: Heuristics for configurational theorizing. Academy of Management Review. 2021; 46(4):778-799. doi: 10.5465/amr.2019.0298.
[32] Lincoln YS, Guba EG. Naturalistic inquiry. Thousand Oaks: Sage Publications; 1985. 416 p.
[33] Zhang L, Wang X, Wang H, et al. Generation pathways of innovation capability among engineering graduate students at Chinese Double First-Class universities: Configurational analysis using the Technology-Organization-Environment framework. Acta Psychologica. 2026 Apr; 264: 106537. doi: 10.1016/j.actpsy.2026.106537.
[34] Liu M, Wang H, Tlili A, et al. Configuring smart education ecosystem for educational quality: A fuzzy-set qualitative comparative analysis of 79 countries. Education and Information Technologies. 2026 Jul. doi: 10.1007/s10639-026-14073-2.
[35] Chen G, Chen P, Wang Y, et al. Research on the development of an effective mechanism of using public online education resource platform: TOE model combined with FS-QCA. Interactive Learning Environments. 2024; 32(10):6096-6120. doi: 10.1080/10494820.2023.2251038.
[36] Geurts EM, Reijs RP, Leenders HH, et al. Promoting student voice by involving vocational education students and staff in citizenship education development: a participatory action research study. Journal of Educational Change. 2025; 26(2):271-289. doi: 10.1007/s10833-024-09523-y.
[37] United Nations Educational, Scientific and Cultural Organization. Global education monitoring report 2023: Technology in education-A tool on whose terms? Global education monitoring report 2023: Technology in education-A tool on whose terms? Paris: UNESCO Publishing; 2023. 526 p.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)

© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261122

Operational Indicator System for School-Enterprise Collaborative Virtual Teaching and Research Offices in Vocational Education: A CIPP Model-Based Study

Huanhuan ZHOU[1]
Guangdong Polytechnic of Science and Trade, Qingyuan, Guangdong 511500, China

Abstract. The school-enterprise collaborative virtual teaching and research office in vocational education integrates digital technology and the modern vocational education system. Based on the CIPP model, this paper constructs an operational indicator system with four dimensions: context, input, process, and product evaluation. The results show that policy‑related factors are closely associated with system operation; resource inputs need to be matched with standardized process implementation to generate expected outputs; talent cultivation and social service constitute core outcome indicators. According to the empirical associations observed in this study, targeted optimization strategies are proposed, including strengthening top‑level policy guidance, building a balanced resource allocation mechanism, and perfecting the process management mechanism for school‑enterprise collaboration. This study provides theoretical and practical reference for the construction of school-enterprise collaborative virtual teaching and research offices in vocational education.

Keywords. School-enterprise collaborative virtual teaching and research office, CIPP model, School-enterprise collaborative education, Structural Equation Model

1. Introduction

In the higher education system of China, the teaching and research office serves as a fundamental grassroots teaching organization and constitutes a core underpinning for the high-quality development of teaching practice. With the continuous diversification of university functions and the steady expansion of institutional scales, traditional teaching and research offices have gradually experienced a decline in their authority and effectiveness within the teaching governance system [1]. Consequently, teaching activities have gradually shifted from organized collective research and practice to individual experience-driven practice, which has exerted a negative impact on the overall teaching quality in higher education institutions.

National policy documents have proposed the national strategy of educational digitalization and promoted AI-enabled teacher development. The digital transformation of vocational education has driven the intelligent upgrading of

[1] Corresponding Author, Huanhuan ZHOU, Guangdong Polytechnic of Science and Trade, Qingyuan, Guangdong 511500, China; E-mail: 25490905@qq.com.

grassroots teaching organizations. Official documents have launched pilot programs for virtual teaching and research offices, and further clarifies the development direction, proposing the construction of school-enterprise virtual teaching and research offices anchored to city-level consortia. This platform can support dual-qualified teachers from higher vocational institutions and enterprises to carry out collaborative teaching and research activities and promote the construction of school-enterprise teaching communities [2].

Following the launch of national pilot projects for virtual teaching and research offices in 2022, academic research in this field has entered a period of rapid expansion. Most of the existing relevant literature were published between 2022 and 2024, indicating that both the academic community and industrial practitioners are still at the early exploration stage of this research area. According to the classification in official pilot documents, virtual teaching and research offices are divided into three categories, which are course-teaching oriented, major-construction oriented, and teaching-reform project oriented. Against this classification framework, considering the core characteristic of school-enterprise integration that is inherent to vocational education, the school-enterprise collaborative virtual teaching and research office forms an independent category that requires targeted special research. This paper focuses on addressing the following three research questions:

(1) What are the core characteristics of the school-Enterprise collaborative virtual teaching and research office in vocational education?
(2) What are the core influencing factors associated with their effective operation?
(3) How do these influencing factors affect the construction and development of the institution?

This paper adopts the CIPP evaluation model as the theoretical framework. Based on questionnaire data from vocational school teachers and enterprise mentors, it constructs a structural equation model to explore the practical logic and operation mechanism of School-Enterprise collaborative virtual teaching and research offices and provides theoretical foundation and practical strategies for their construction in higher vocational education.

2. Literature Review

2.1 Virtual Teaching and Research Office

There are currently three main academic interpretations of virtual teaching and research offices. First, a temporary matrix organization dissolved after project completion. Second, an online interactive platform for teachers and students. Third, an emerging grassroots teacher organization relying on new-generation information technology for cross-school and cross-disciplinary collaborative teaching research [3]. The third definition is consistent with official pilot documents, and this study adopts it as the analytical scope.

Existing studies mainly concentrate on practical experience summarization, characteristic analysis, and construction path exploration [4]. Most relevant achievements focus on general undergraduate education, and research in vocational education is still emerging [5]. Existing scholars have conducted targeted research in subdivided fields such as entrepreneurship education and college-articulation programs,

and the academic community has recognized its value in promoting multi-stakeholder collaboration in vocational education.

Vocational education has been plagued by persistent pain points: passive enterprise participation, fragmented school-enterprise collaborative, formalized practical teaching, and mismatch between talent cultivation and industrial demand. The root cause is insufficient depth of school-enterprise collaborative. Existing reform modes including industrial colleges, modern apprenticeship and order-based talent cultivation target constructing a sustainable school-enterprise interest community. To maintain stable operation of collaborative mechanisms, institutionalized grassroots organizations equipped with standardized communication channels, professional platforms and operation rules are necessary. The school-enterprise integrated virtual teaching and research office meets these requirements, and its emergence is an inevitable outcome when vocational education and industry integration enters the stage of high-quality digital development.

2.2 Operational Indicators of School-Enterprise Virtual Teaching and Research Offices in Vocational Education Based on the CIPP Model

The CIPP evaluation model was proposed by American scholar Daniel L. Stufflebeam as an improvement to Tyler's behavioral objective evaluation framework [6]. Unlike Tyler's model, which focuses on predetermined objectives and quantifiable behavioral outputs, it overlooks the holistic and procedural characteristics of educational activities. The CIPP model divides project operation into four core dimensions: context, input, process and product evaluation. It enables comprehensive evaluation of organizations and projects under dynamic conditions and provides guidance for continuous quality improvement. After decades of revision, the CIPP model has been widely applied in policy evaluation and educational project optimization. International research has verified its applicability: Yastibas et al. applied it to academic English teaching in Turkish universities [7]; Molope et al. evaluated community development programs in South Africa [8]; Pramono et al. assessed school-based sports programs in Indonesia [9]. In TVET, researchers have used the CIPP model to evaluate industry matching programs of vocational high schools [10]. Studies on virtual teacher collaboration platforms confirm that process indicators like interaction quality and collaborative task completion rate significantly predict teachers' professional growth [11]. These international achievements provide theoretical and methodological references for this study.

2.2.1 Context Evaluation

Context evaluation analyzes the internal and external environments of the project, identifies the needs of the target group, development opportunities, and practical constraints, and verifies whether project objectives respond to real-world demands. This study divides the context dimension into three analytical levels: environmental context, problem context, and resource context. From the external perspective, the national digital education strategy and industry-education integration reform have put forward development demands for the construction of virtual teaching and research offices [12]. From the internal perspective, traditional discipline-based bureaucratic teaching organizations have shown limitations in adapting to new development requirements [2]. The main obstacles restricting the construction of school-enterprise

collaborative virtual teaching and research offices include insufficient inter-school collaborative foundation, low efficiency of school-enterprise information sharing, insufficient deep integration of digital technology and vocational education teaching, and the generally low digital literacy of school teachers and enterprise tutors [13]. Meanwhile, the continuous construction of dual-qualified innovative teacher teams and the widespread popularization of digital education tools have provided necessary human and material support for the development of virtual teaching and research offices. (Table 1)

Table 1. Context Evaluation Indicators for School-Enterprise Virtual Teaching and Research Office in Vocational Education

First-Level Indicator	Second-Level Indicator
Environmental Context	External Environment
	Internal Environment
Problem Context	Chronic Problems of School-Enterprise Integration
	Digital Application Problems
	Traditional Teaching and Research Office Construction Problems
Resource Context	Dual-Qualification Team Building
	Popularization of Teaching Information Technology Application
	Improvement of Teachers' Digital Literacy

2.2.2 Input Evaluation

Input evaluation focuses on the sufficiency and rationality of resources invested in project implementation, covering four dimensions: capital, human resources, physical infrastructure and policy planning. Reasonable resource input is the fundamental prerequisite for smooth implementation.

Table 2. Input Evaluation Indicators for School-Enterprise Virtual Teaching and Research Office in Vocational Education

First-Level Indicator	Second-Level Indicator
Financial Input	Special Funds
Human Input	Core Team Building
	Management Department Building
	Training and Cultivation
Material Input	Platform Construction
Policy Input	School Support Policy Issuance
	School-Enterprise Virtual Teaching and Research Office Operation Plan Formulation

Financial input includes three core elements: funding source, investment scale and standardized transparent financial management system. Human input involves organizational structure, core team building, role responsibility definition and digital literacy improvement of teachers from vocational schools and cooperative enterprises [10]. Material input centers on information platform construction as operational carriers for collaborative teaching research, resource sharing and teacher training. Policy input focuses on top-level design, special development planning and standardized institutional rules. Based on the above analysis, this research sets four first-level indicators: financial input, human input, material input and policy input. (Table 2)

2.2.3 Process Evaluation

Process evaluation focuses on dynamic monitoring of the project implementation process, identifying and addressing real-time problems in implementation, to ensure

that all related activities can advance in accordance with predetermined objectives. For vocational education School-Enterprise Virtual Teaching and Research Office, their core practice tasks cover teaching resource development, talent cultivation implementation, teacher professional training, social service delivery, and teaching research reform [10]. Different from conventional virtual teaching and research offices, vocational education-oriented virtual teaching and research offices pay higher attention to school-enterprise collaborative modular teaching reform, on-site practical teaching and industry-oriented social service functions. Based on this, this study constructs five first-level process indicators accordingly. (Table 3)

Table 3. Process Evaluation Indicators for School-Enterprise Virtual Teaching and Research Office in Vocational Education

First-Level Indicator	Second-Level Indicator
Teaching Resource Construction	Teaching Material Construction
	Course Construction
	Practical Teaching Resource Construction
Talent Cultivation	Classroom Teaching
	On-Site Work Teaching
Teacher Training	School Teacher On-Site Teaching Ability Cultivation
	Enterprise Mentor On-Site Teaching Ability Cultivation
Social Service	Social Service

2.2.4 Product Evaluation

Product evaluation measures the level of goal achievement, the overall influence, actual effectiveness and long-term sustainability of the project. The sustainable operation of virtual teaching and research offices relies on the formation of a consistent interest connection between schools and enterprises. The evaluation of project outcomes is divided into three dimensions: school-side effectiveness, enterprise-side effectiveness and overall industry-education integration benefits. It is important to note that industry-education integration operates at the macro policy level, and provides the policy environment. From the school's perspective, the core objectives of the project cover two key aspects: talent cultivation quality (including indicators such as graduate employment matching rate and employment quality) and the professional growth of teachers. From the enterprise's perspective, the expected benefits they obtain from participating in the construction include reduced recruitment costs, access to intellectual support and technical problem-solving services provided by school teachers. The realization of mutually beneficial sharing is the foundation for building a stable school-enterprise interest community. Accordingly, this research sets four first-level product evaluation indicators: talent cultivation quality, social service effectiveness, teacher professional growth, and deep industry-education integration. (Table 4)

Table 4. Product Evaluation Indicators for School-Enterprise Virtual Teaching and Research Office in Vocational Education

First-Level Indicator	Second-Level Indicator
Talent Cultivation Quality	Major Match Rate
	Graduate Enterprise Retention Rate
Social Service	Helping Enterprises Improve Production Efficiency
Teacher Growth	School Teacher On-Site Teaching Ability Improvement
	Enterprise Mentor On-Site Teaching Ability
Deep Industry-Education Integration	School-Enterprise Integration Community

3. Research Design

3.1 Data Sources

Based on the CIPP theoretical framework, this study designed a structured questionnaire covering context evaluation, input evaluation, process evaluation and product evaluation of four dimensions. The survey respondents were restricted to school teachers and enterprise mentors who have more than one year of working experience in school-enterprise collaborative education. A total of 130 questionnaires were distributed and recovered, and 126 valid samples were obtained after screening, with an effective response rate of 96.9%. The composition of valid samples is as follows: school teachers account for 56.4%, and enterprise mentors account for 43.6%; most respondents are ordinary employees or grass-roots managers. The overall Cronbach's α coefficient of the questionnaire is 0.972, indicating that the scale has good internal consistency reliability. After item purification based on discriminant validity test, 21 formal measurement items are finally retained. During discriminant-validity testing, nine initial items were eliminated. The deletion criteria were formulated as follows: (1) factor loading lower than 0.50; (2) items with cross-loading values across two or more latent variables differing by less than 0.20; (3) items whose content semantics deviate from the connotation of their intended latent variable (the so-called "misplacement"). Each eliminated item was checked against theoretical definitions of CIPP dimensions, and retained items demonstrate clear conceptual correspondence to their respective latent construction. After item purification, 21 formal measurement items are retained. KMO values for all four latent variables exceed 0.6, satisfying prerequisites for factor analysis.

3.2 Structural Equation Model Analysis

In this study, context evaluation is set as the exogenous variable, while input evaluation and process evaluation are set as mediating variables, and product evaluation is set as the endogenous variable. The model fitting test results show that: $\chi^2/df = 2.178$, GFI = 0.992, CFI = 0.974, AGFI = 0.985, RMSEA = 0.020. All fitting indices meet the model acceptance criteria, and all path coefficients reach the statistically significant level ($P<0.05$).

The quantitative analysis results indicate that the path coefficients of the three dimensions under context evaluation are as follows: "environmental context" 0.459, "problem context" 0.390, and "resource context" 0.151. Among the input dimensions, the path coefficient of financial input is 0.294, which ranks first, followed by policy input (0.274), material input (0.269), and human input (0.267). Within the process evaluation dimension, social service (0.410) and teacher training (0.302) have the highest impact weights. For the product evaluation dimension, social service outcomes (0.373) and talent cultivation quality (0.355) are the core outcome indicators.

3.3 Mediation Effect Test

The Bootstrap method was used to conduct the mediation effect test (5000 resamples, 95% confidence interval). The test results show that process evaluation plays a partial mediating role between input evaluation and product evaluation. In this path, the

indirect effect value is 0.625 (95% CI [0.498, 0.752], *p* < 0.001), and the direct effect value is 0.596 (95% CI [0.473, 0.719], *p* < 0.001). The mediating effect accounts for 51.2% of the total effect. This finding suggests that resource input is largely associated with final outcomes through process execution. Context evaluation can affect the final outcome not only through a direct path, but also through sequential multiple mediation chains.

4. Research Results

The regression results show that environmental context (β =0.459) and policy input (β =0.274) have strong statistical associations with outcomes. The operation of school-enterprise collaborative virtual teaching and research offices is closely related to macro-level policy conditions. National pilot policies and school-level institutional documents clarify the rights and obligations of cooperative schools and enterprises, effectively eliminating role ambiguity between participating subjects. This empirical finding is consistent with the core viewpoint of institutional theory: macro policies can impose isomorphic pressure on grass-roots vocational education organizations [14]. Different from the market-driven education operation system, in the research scenario of this study, the influence of policy context is significantly greater than that of resource context.

Human, financial, material and policy inputs should be allocated in a balanced manner, among which special fund allocation and fund utilization efficiency constitute critical guarantee conditions. Given that the indirect effect (0.625) is greater than the direct effect (0.596), resource input cannot automatically produce expected output. The value realization of resources depends on the organizational process of resource allocation and integration [15]. More than half of the input-output transformation relies on process management, which is consistent with the sequential logic proposed by the CIPP evaluation framework. However, given the cross-sectional nature of the data, these findings indicate correlational patterns rather than definitive causal relationships. These findings support the associative path logic of the CIPP evaluation model rather than confirming deterministic causal mechanisms.

The path coefficients of talent-cultivation quality and social-service outcomes are 0.373 and 0.355, respectively. Based on stakeholder theory, vocational schools prioritize the achievement of talent-cultivation goals, while cooperative enterprises have demands for technical support and economic returns [16]. For the sustainable development of virtual teaching and research offices, the organization must realize value co-creation for both stakeholders. This indicates that the product evaluation system under the multi-stakeholder cooperative mode must integrate the dual benefits of the education sector and the industrial sector.

Social service (β =0.41) and school-enterprise collaborative teacher training (β =0.302) are the core process tasks, with path coefficients higher than those of talent cultivation (β =0.222) and teaching resource construction (β =0.191). According to organizational lifecycle theory, newly established collaborative organizations tend to prioritize short-term tangible benefits to consolidate cooperative partnership and build mutual trust among all parties, before advancing the long-cycle talent cultivation reform [17]. Capacity building for dual-qualification faculty is the fundamental support for the operation of the organization: social service can effectively meet the realistic

interest demands of cooperative enterprises, while teacher training can address the problem of insufficient cross-boundary teaching competence for both in-school teachers and enterprise mentors.

This study extends the application of the CIPP model from general program evaluation to the specific context of school-enterprise collaborative organizations in vocational education. By operationalizing the four CIPP dimensions within the virtual teaching and research offices setting, the findings reveal that the relative importance of evaluation dimensions may vary across organizational types: in school-enterprise collaborative contexts, policy context exerts stronger influence than resource context, and process indicators related to social service and teacher training carry greater weight than traditional teaching resource construction. These findings refine the CIPP framework by identifying context-specific indicator weights and by demonstrating that the product evaluation dimension must incorporate dual-stakeholder outcomes rather than single-sector metrics. Furthermore, the identification of process evaluation as a critical mediator between input and product dimensions provides empirical support for the procedural emphasis of the CIPP model in collaborative organizational settings.

5. Policy Recommendations

The empirical results indicate that macro policies and internal school policy inputs significantly impact the construction effectiveness of school-enterprise collaborative virtual teaching and research offices. The government should formulate special guiding documents conforming to vocational education characteristics, establish a targeted policy system, and increase support for school‑enterprise collaborative education. Vocational colleges should issue localized implementation documents clarifying competent management departments, specific objectives, core tasks, guarantee measures, rights and responsibilities of schools and enterprises, and evaluation standards. Each specialty can carry out construction activities based on existing industry‑education integration communities, industrial colleges and modern apprenticeship platforms.

In the initial exploration stage, the balanced allocation of human, financial, material and policy resources are the core prerequisite for development. It is necessary to set up special construction funds and establish a transparent and standardized fund management system. Organize a clear core work team and carry out special digital skill training for school and enterprise teachers participating in the platform. Complete the construction of high‑level information infrastructure and professional virtual teaching and research platforms. Schools should also improve supporting institutional documents and work operation plans to provide a standardized institutional guarantee for the operation of the organization.

Social service is both a core process task and a key output outcome. It is necessary to fully leverage the platform advantages of school-enterprise communication to reduce information asymmetry between participating subjects. Carrying out targeted enterprise service projects can effectively address practical problems encountered in enterprise production and promote technological innovation and benefit improvement of enterprises. Meanwhile, it is essential to establish a standardized service feedback mechanism to systematically collect satisfaction data from cooperative enterprises. Evaluation indicators should cover service scale, service quality, enterprise satisfaction

and industrial contribution, and the evaluation results should be used as an important basis for the iterative optimization of the school-enterprise collaborative virtual teaching and research office.

Mediation analysis reveals that resource inputs relate to project outcomes through effective process implementation. Schools and enterprises should jointly develop detailed process management schemes, clarify task content, schedule and responsible subjects, and establish a regular supervision mechanism. First, a competency-oriented teaching resource system should be constructed, covering co-developed materials, courses, projects and databases, with curriculum adjusted from discipline-oriented to job requirement orientation. Second, a scientific training scheme for teachers and mentors should be designed to improve practical teaching ability. Training effect should be linked with title evaluation and performance appraisal. Third, standardized on-site practical teaching should be promoted based on real production scenarios, realizing teaching standardization, clarifying responsibilities, and breaking the boundary between campus and enterprise training.

6. Limitations

This study still has three primary limitations. First, the data adopted in this study are cross-sectional data collected at a single time point. Although structural equation modeling (SEM) and mediation analysis can identify correlation patterns that are consistent with the proposed CIPP causal chain, cross-sectional data cannot effectively verify causal relationships. Future research can adopt longitudinal research design or quasi-experimental design to track the developmental process of the research object and further test the causal direction of relevant relationships.

Second, the research sample is limited to Guangdong Province, which is one of the most economically developed regions in China. Guangdong benefits from advanced digital infrastructure, mature industry-education integration policies, and substantial fiscal support for vocational education−conditions that may not be representative of less developed regions. Therefore, the research conclusions may not be generalized to regions with different levels of resource endowment, policy support and school-enterprise collaboration maturity. For instance, western and central provinces may face more severe constraints in enterprise participation and digital infrastructure, which could alter the relative importance of context and input indicators identified in this study. Future multi-regional comparative studies can effectively improve the external validity of the research conclusions.

References

[1] Hong Zhizhong. Evolution and Reconstruction of Grassroots Teaching and Research Offices in Universities. University Education Science, 2016(3): 86-92.

[2] Bie Dunrong. Functions and Construction Paths of Virtual Teaching and Research Offices. China Higher Education Research, 2024(4): 7-14.

[3] Zhan Dechen, Nie Lanshun, Tang Dekai, et al. Virtual Teaching and Research Office: A New Form of Collaborative Teaching and Research. Modern Educational Technology, 2022(3): 23-31.

[4] Yan Xiuying, Yu Junqi, Feng Zengxi, et al. Exploration Research on Construction Path of Building Electrical and Intelligent Virtual Teaching and Research Office. Education and Teaching Forum, 2024(45): 117-120.

[5] Chen Jing, Yang Yinyu, Deng Fengjun, et al. Exploration of Virtual Teaching and Research Office Construction from the Perspective of school-enterprise Integration Community. Office Automation, 2024(18): 90-93.
[6] Kellaghan T, Stufflebeam D L. The CIPP Model for Evaluation. Springer Netherlands, 2003, 10(4): 31-62.
[7] Yastba A E, Erdal T. Evaluating English for Academic Purposes II Course through the CIPP Model. Gumushane University Electronic Journal of the Institute of Social Science, 2020, 11(1): 86-94.
[8] Molope M, Oduaran A. Evaluation of the Community Development Practitioners' Professional Development Programme: CIPP Model Application. Development in Practice, 2020, 30(2): 194-206.
[9] Pramono R, Sarliyani S, Purwanto A. The Evaluation of Narada Cup School Sport Program Using CIPP Evaluation Model. Jurnal Pendidikan Jasmani Dan Olahraga, 2020, 5(1): 81-86.
[10] Sun Jie, Wu Zhaoming, Liu Fahu, et al. Value Appeals and Construction Paths of Virtual Teaching and Research Offices in Higher Vocational Colleges under the Background of Educational Digital Transformation. Chinese Vocational and Technical Education, 2023(31): 74-79+87.
[11] Su Xiaohong, Miao Qiguang, Chen Wenyu. Personalized Teaching Model for Improving Programming Design Ability Based on AI Empowerment and school-enterprise Integration. China University Teaching, 2023(6): 4-9.
[12] Xue Quan. Construction of Vocational Education Teacher Innovation Teams under Duality Theory: Theoretical Research, Operation Mechanism, and Prudent Considerations. Vocational Education Forum, 2024(12): 67-73.
[13] Shen Zhongyan, Sun Dan. Construction of Teacher Teaching Innovation Teams in Vocational Colleges: Logic, Effectiveness, and Future Directions: Based on Research of 32 Teams. Vocational and Technical Education, 2025(2): 54-60.
[14] DiMaggio P J, Powell W W. The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. American Sociological Review, 1983, 48(2): 147-160.
[15] Barney J. Firm Resources and Sustained Competitive Advantage. Journal of Management, 1991, 17(1): 99-120.
[16] Freeman R E. Strategic Management: A Stakeholder Approach [M]. Boston: Pitman, 1984.
[17] Lunnan R, Haugland S A. Predicting and Measuring Collaboration Outcomes in Alliances: A Study of the Norwegian Seafood Industry. Journal of Business Research, 2008, 61(6): 648-658.

Section 4

Special Session on “The Potential of Big Data, Digitalization and AI in a Sustainable Economy”

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261124

AI-Driven Export Packaging Design: Challenges of Cultural Differences and Intelligent Response Strategies

Aihua Song[1]
School of Digital Economics, Changzhou College of Information Technology, Changzhou, China, 213164

Abstract. Cultural differences profoundly affect the international market acceptance of products in global trade. Packaging, as the first point of contact with consumers, undertakes an important mission of cross-cultural communication. This article starts from the perspective of intelligent design, systematically analyzes the impact and risks of cultural differences in dimensions such as color, graphics, language and text, religious beliefs, and consumption habits on packaging design, and proposes an intelligent packaging design solution supported by AI technology, driven by data, and centered on dynamic adaptation. By constructing an AI-enhanced design process of "research-generation-testing-optimization", the aim is to provide efficient, accurate, and iterative cross-cultural packaging strategies for Chinese export enterprises, reduce the risk of cultural conflicts, and enhance brand international competitiveness and market response speed.

Keywords. Cultural difference; export packaging; artificial intelligence; cross-cultural design; intelligent localization

1. Introduction

In the fierce international market competition, packaging has evolved from a simple product protection container to a key marketing tool that conveys brand value and influences purchasing decisions. However, many Chinese export enterprises, despite producing high-quality products, have encountered "cultural mismatches" in packaging design due to insufficient understanding of the cultural characteristics of the target market, which has hindered market expansion [1]. Resolving cultural differences requires a structured understanding of the culture itself. According to Hofstede's theory of cultural dimensions and Hall's theory of high and low context culture, culture is shaped into a multi-level system that includes tangible and intangible elements of perception and behavior, encoded into analyzable data points such as color semantics, symbol popularity, language patterns, and value indicators extracted from multimodal sources. With the rapid development of artificial intelligence technology, packaging design is entering a new stage of intelligence and data-driven development. AI can not only efficiently

[1] Corresponding Author: Aihua Song, School of Digital Economics, Changzhou College of Information Technology, Changzhou, China; Email: sah1001@126.com

identify and analyze design elements in cross-cultural contexts, but also achieve rapid localization and adaptation of packaging through generative design, emotional computing, and behavioral prediction. Therefore, exploring the application of AI in addressing cultural differences and building an intelligent packaging design system has important theoretical value and practical significance.

2. Intelligent recognition and analysis mechanism for cultural differences

Culture is a complex system that influences the entire process of packaging design. Artificial intelligence can systematically identify and quantify the expression of cultural differences in design through multimodal data perception and cultural semantic analysis. The technical pathway involves a multi-stage pipeline.

(1) Data Ingestion & Feature Extraction: Collecting multimodal data (images, text, sales data) from target markets. Computer Vision (CV) and Natural Language Processing (NLP) models extract low-level features (color histograms, object labels, sentiment scores, keyword frequencies).

(2) Cultural Encoding & Mapping: These features are mapped to pre-defined cultural dimensions and knowledge graphs (e.g., a "Cultural Symbol Database" linking "owl" to "wisdom" in Western contexts vs "misfortune" in some Eastern ones).

(3) Modeling & Inference: Machine learning models (e.g., classifiers, GANs) are trained on this encoded data to learn associations between cultural features and market performance or user preference metrics.

(4) Decision Rule Application: Rule-based systems or optimization algorithms translate model inferences into concrete design parameters (e.g., "IF market=Japan AND product=tea, THEN recommend color palette with saturation < X, incorporate minimalist patterns").

The impact mechanism can be represented by Figure 1:

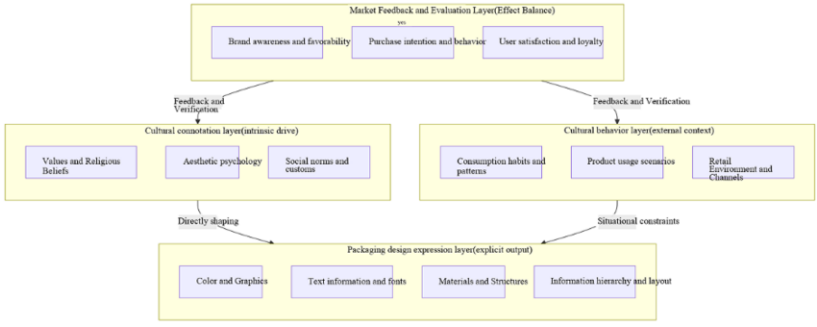


Figure 1. Model of the impact mechanism of cultural differences on packaging design

Figure 1 illustrates how culture systematically influences various performance elements of packaging design, from core values to behavioral consumption habits, and ultimately affects market outcomes. Specifically, its impact is mainly reflected in the following dimensions:

2.1 Intelligent recognition of color preferences and taboos

Color is the most intuitive cultural carrier in packaging, and its symbolic meaning varies with different cultures. AI can quickly construct a "global color culture map" through image recognition and sentiment analysis technology, identifying color preferences and taboos in different regions.

(1) AI driven color semantic analysis. Utilize computer vision to analyze the frequency of color usage and emotional feedback in social media, e-commerce platforms, and advertising materials, and automatically establish a regional color database. For example, AI can recognize the positive meanings of red in Chinese and Indian festivals, while warning of its negative associations in some African countries.

(2) Dynamic color scheme generation. Based on Generative Adversarial Networks (GANs) trained on culturally-tagged image datasets, AI can generate multiple color matching schemes based on cultural data of the target market and predict their market acceptance. For example, for the Japanese market, a cool color scheme that embodies "refinement and technology" is generated, while for the Southeast Asian market, a high saturation warm color combination that embodies "passion and vitality" is generated.

2.2 Intelligent screening of graphics, symbols, and patterns

The cultural interpretation of graphics and symbols differs significantly, and AI can perform automated symbol recognition and cultural sensitivity detection.

(1) Cultural symbol database and risk warning. Establish a database covering the meanings of multicultural symbol, combined with natural language processing (NLP) and image recognition, to automatically detect patterns in design drafts that may cause misinterpretation or aversion (such as dragons, owls, specific geometric shapes), and provide replacement suggestions.

(2) Intelligent pattern generation and adaptation. By utilizing style transfer techniques and cultural feature extraction, AI can automatically adapt brand core graphics to the aesthetic style of the target market. For example, the traditional patterns of Chinese tea brands are automatically transformed into design variants that conform to Japanese Zen aesthetics or British style.

2.3 Intelligent localization of language and text

The translation and layout of language and text are key to cross-cultural packaging [2-7], and AI has significant advantages in improving translation accuracy and maintaining brand tone.

(1) AI assisted translation and context adaptation. By combining neural machine translation with a terminology library and context-aware models fine-tuned on marketing corpora, brand copy can be quickly and accurately translated, and automatically adapted to cultural contexts. For example, avoid the literal translation of "Baijiu" into "white wine", but use expressions such as "Baijiu (Chinese Spirits)" that conform to local cognition.

(2) Intelligent fonts and layout optimization. AI can analyze the font usage and layout of mainstream packaging in the target market, recommend the most readable and emotionally compatible font, and automatically adjust the layout direction (such as for Arabic reading habits from right to left).

2.4 Intelligent compliance of religion and social values

Religious beliefs and social values directly affect consumers' purchasing decisions, and AI can assist companies in achieving certification label management and aligning values.

(1) Automatic recognition and embedding of certification marks. Through image recognition, automatically detect whether the packaging contains necessary religious certification marks (such as Halal, Kosher) and ensure that their position and size comply with specifications.

(2) Intelligent embodiment of sustainability values. AI can analyze the preferences of target market consumers for environmentally friendly materials and minimalist designs [8], recommend packaging materials that meet local green standards, and automatically generate environmental declaration texts.

2.5 Data driven design of consumption habits and usage scenarios

The physical properties of packaging need to match the local lifestyle, and AI can optimize packaging structure, size, and functional design through consumer behavior data.

(1) Optimization of packaging specifications based on purchase data. By utilizing sales data and user reviews, AI can recommend the most popular product sizes and packaging forms, such as small portion packaging in the European and American markets and large capacity home packaging in the Middle East market.

(2) Retail environment simulation and structural testing. By combining virtual reality (VR) and augmented reality (AR) technologies, AI can simulate the display effect of packaging on target market shelves and test its structural strength and logistics adaptability.

Conceptual Framework Integration: The mechanisms described above are unified under a proposed AI-Driven Cross-Cultural Packaging Design (AI-CCPD) Framework. This framework posits that effective cross-cultural adaptation is achieved through a continuous loop of: (A) Cultural Intelligence Acquisition (data collection & encoding), (B) Computational Cultural Modeling (inference & rule generation), and (C) Design Synthesis & Validation (generation & testing). AI serves as the engine connecting (A) to (C), transforming abstract cultural dimensions into concrete, optimized design parameters.

3. AI-driven cross-cultural packaging design system process

Based on the above intelligent recognition mechanism, enterprises should establish a cross-cultural packaging design process that is "data-driven, AI enhanced, and human-machine collaborative", as shown in Figure 2:

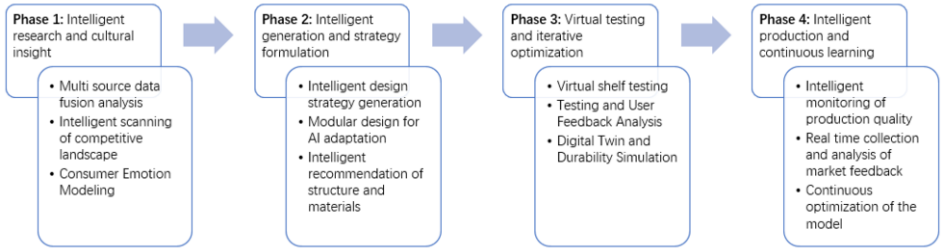


Figure 2. Flow chart of AI-driven cross-cultural packaging design system

This process deeply integrates traditional design processes with AI technology to form a closed-loop iterative intelligent design system. Its novelty lies in the pervasive, data-driven AI automation at each stage, moving beyond mere digital assistance to predictive generation and autonomous optimization based on cultural models.

3.1 Phase 1: Intelligent research and cultural insight

AI can replace or assist manual labor in completing large-scale, multi-dimensional market data analysis, improving research efficiency and depth. This phase is predominantly AI-automated.

(1) Multi source data fusion analysis. Integrate multi-channel data from social media, e-commerce platforms, industry reports, etc., and use NLP and image recognition technology to automatically extract cultural trends, color preferences, symbol meanings, and other information.

(2) Intelligent scanning of competitive landscape. AI automatically captures images and copy of competitor packaging, compares and analyzes visual elements, material processes, and information layout, and identifies market opportunities and design gaps.

(3) Consumer sentiment modeling. Quantify consumers' emotional feedback on existing packaging through sentiment calculation and opinion mining, and predict their acceptance of different design elements.

The output of this stage is the 'AI Culture Insight Report', which provides data support for designing strategies.

3.2 Phase 2: Intelligent generation and strategy formulation

Based on research data, AI assisted or self-directed generation and optimization of design solutions. This phase is AI-driven with human curation.

(1) Intelligent design strategy generation. AI automatically generates design guidelines based on cultural insights, including color combinations, graphic styles, font suggestions, information hierarchy, etc., to ensure that the design direction conforms to the cultural tone of the target market [9].

(2) Modular design for AI adaptation. On the basis of fixed core elements of the brand, AI automatically adjusts variable modules (such as graphics, colors, and layouts) to quickly generate multiple localized design versions [10], achieving efficient output of "one product with multiple appearances".

(3) Intelligent recommendation of structure and materials. Combining logistics data with environmental regulations, AI recommends the most suitable packaging structure

and material solutions, such as moisture-proof and seismic design, environmentally friendly and recyclable materials, etc.

This stage can output multiple sets of packaging design solutions generated or optimized with AI assistance.

3.3 Phase 3: Virtual testing and iterative optimization

Utilize digital tools for low-cost and efficient design validation. This phase is data-driven and AI-analytic.

(1) Virtual shelf testing. Simulate real retail scenarios through VR technology to evaluate the visual prominence and attractiveness of packaging among competitors. AI algorithms analyze eye-tracking simulation data to score visibility.

(2) Testing and user feedback analysis. Publish different design versions online, collect user interaction data, and use AI to analyze preference trends in real-time to guide design iterations.

(3) Digital twin and durability simulation. By using simulation technology to test the physical performance of packaging during transportation, stacking, and storage, structural defects can be detected in advance.

At this stage, design optimization suggestions are formed and fed back to the generation stage to achieve rapid iteration.

3.4 Phase 4: Intelligent production and continuous learning

After the design is finalized, the AI system continues to learn and evolve in the market. This phase enables system evolution.

(1) Intelligent monitoring of production quality. Utilize computer vision to inspect the printing and finished product quality on the production line, ensuring consistency with the design draft.

(2) Real time collection and analysis of market feedback. Continuously monitoring sales data, user reviews, and social media volume, AI automatically summarizes the market performance and cultural acceptance of packaging.

(3) Continuous optimization of the model. Based on market feedback data, update the cultural parameters and design rules in the AI model to make it more accurate [11], providing intelligent support for subsequent product iterations or new market entry.

4. Illustrative case: AI-CCPD system application for a Chinese green tea brand entering the UK and Saudi Arabian markets

To verify the feasibility of the proposed system, a simplified example is provided. Input: brand core assets (logo, Chinese-style pattern), target markets (UK, Saudi Arabia). Phase 1: AI analysis reveals the UK market's preference for minimalist design, traditional elements (such as subtle patterns), and environmental protection information; the Saudi market prefers gold/black, exquisite patterns, and sizes suitable for family sharing. Phase 2: AI generates two distinct modular designs: for the UK market, a stylized version of the core pattern using a white/green color palette, emphasizing "handmade" and "sustainable" copywriting. For the Saudi Arabian market, a gold/black/green color palette, elaborately decorated pattern frame, and copywriting emphasizing "high-end

gift" and "family gathering" are adopted. Phase 3: Virtual shelf testing shows that the UK design integrates well in health food stores; the Saudi design stands out in the luxury gift area. A/B testing confirms preferences. Output: two culturally adapted packaging versions, along with feedback performance data, to enrich the AI model for future projects.

5. Conclusion and Prospect

In today's world of globalization and digitization, export packaging design is shifting from "experience driven" to "data-driven" and from "static adaptation" to "dynamic intelligence". Artificial intelligence provides a new technological path and strategic framework for cross-cultural packaging design through efficient cultural recognition, intelligent content generation, and precise testing feedback. Enterprises should actively embrace AI tools, build a human-machine collaborative design system, and achieve precise adaptation and rapid response of packaging in the cultural dimension.

In the future, with the further development of technologies such as multimodal big models, generative AI, and virtual reality, cross-cultural packaging design will become more automated, personalized, and contextualized. Enterprises can not only prevent cultural risks through AI, but also actively explore cultural resonance points, achieve deeper emotional connections between brands and consumers, and establish lasting competitiveness in the global market.

References

[1] Yang H. Research on Export Packaging Design of Cross border E-commerce under Cross Cultural Background. China Packaging, 2022; 42(08):46-48.
[2] Peng Y.F. Exploration of English Translation Practice for Text on Export Food Packaging. Journal of Nuclear Agricultural Sciences, 2021; 35 (11): 2686-2687.
[3] Zhang X. English Translation Strategies for Food Packaging in Export Trade. China Food, 2023; (24):123-125.
[4] Di F.J. Attribution Analysis and Improvement Strategy of Translation of Exported Food Packaging Texts. Journal of Food Safety & Quality, 2019; 10(18):6369-6375.
[5] Wang K. Strategies for Improving English Translation of Tea Packaging for Export in the Internet Environment. Tea in Fujian, 2024; 46 (01): 158-160.
[6] Lu W. Research on the English Translation of Packaging Text for "Gan Flavor" Agricultural Products in the Context of the Belt and Road Initiative. China Packaging, 2023; 43(05):9-14.
[7] Li K.N., Ni P. Chinese Character Style Design of Alphabetic Characters on Exported Tea Package from the Perspective of the Cha-Ma Ancient Route. Packaging Engineering, 2022; 43(18):434-440.
[8] Wu H.B., Zhang Y. The Impact of Green Barriers in Packaging on the Export of Enterprises. Journal of Anhui Agricultural University (Social Sciences Edition), 2021; 30(04):67-71.
[9] Guo J.Z., Mao H. Visual Design Strategies for Cross-Border Food Packaging Based on Acculturation Theory. Food & Machinery, 2024; 40(12):204-209.
[10] Yang M.J. Localization Tendency and Correction of Chinese Cultural Characteristics in Export Food Packaging. Food & Machinery, 2021; 37(12):110-112.
[11] Faraghian H, Dolatabadi H R, Esmaelian M. Model Design for Factors Affecting the Packaging of Iranian Fruit and Vegetables for Export. International Journal of Services and Operations Management, 2022; 42(3):339-352.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)

© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261125

The Effects of Corporate Digital Transformation on Financial Information Governance: Evidence from Financial Restatements

Zhe Zhang[a] and Ximeng Jia[b,1]
[a] *CCCC Infrastructure Maintenance Group Co., Ltd., General Contracting Branch, China*
[b] *Business School, Sichuan University, China*

Abstract. Digital transformation presents opportunities for corporate governance, particularly in financial information management. This study investigates how digital transformation shapes financial information governance, focusing on financial restatements. Based on empirical analysis, this paper demonstrates that digital transformation significantly reduces the occurrence of financial restatements, thus improving financial reporting quality. Robustness checks confirm the reliability of these findings. Mechanism analysis suggests that digital transformation impacts financial restatements by reducing information asymmetry, lowering agency costs, and optimizing internal controls. Further analysis reveals that the effectiveness of digital transformation is diminished in firms with CEO-chair duality, financing constraints, and state-owned enterprises. This study provides new evidence on the role of digital empowerment in corporate governance and offers insights for the development of digital economies and government regulatory policies.

Keywords. Digital transformation, financial restatement, financial information government, digital technology

1. Introduction

Amid rapid digital advances, enterprises increasingly embrace digital transformation to restructure organizations and upgrade business models [1]. Driven by digital technologies, firms can produce and manage more efficiently, expand and innovate their businesses, and better meet customer demands [2]. Meanwhile, with growing market competition, enterprises need to continuously enhance their core competitiveness, which can be achieved through digital transformation to innovate and optimize business processes, strengthen brand value, and increase market share. However, some studies have found that investment in digitalization may yield minimal returns or even reduce corporate performance, a phenomenon referred to as the "digital paradox" [3]. Whether enterprises can improve management efficiency and maintain stable and healthy market and social development through digital transformation remains an important theoretical and practical issue.

Digital transformation often involves the use of various digital technologies, such as data management systems and machine learning, which can reshape financial information collection, processing, and reporting [4]. Firms may leverage digital technologies to streamline business

[1] Corresponding Author: Ximeng Jia, jiaxm@scu.edu.cn.

processes and embed activities into standardized systems, reducing human discretion or ambiguity in financial information generation [5]. In a digital context, managerial opportunism is constrained not only by formal governance mechanisms but also by technology-driven process controls. Digital technologies, such as enterprise resource planning (ERP) systems, automated accounting platforms, and data-driven internal control systems, embed standardized rules and real-time monitoring into firms' financial reporting processes. By limiting manual adjustments, enhancing transaction traceability, and increasing the transparency of accounting procedures, these technologies significantly reduce managerial discretion. Furthermore, new business models and transaction methods associated with digitalization may increase the complexity and difficulty of financial information processing [6].

Financial information is a key channel through which firms communicate their operating performance to external stakeholders, playing a critical role in corporate management, investor decision-making, and capital market stability [7]. High-quality financial information enhances corporate transparency, clarifies operational and financial conditions, and allows potential investors to assess the credibility and attractiveness of the firm, thereby increasing investment opportunities [8]. However, managers may engage in selective disclosure, manipulate financial data, or fabricate business activities to smooth earnings, boost compensation, or meet analysts' forecasts, reducing the quality of financial information [9]. Financial restatements, as corrections to previously issued financial statements, are widely recognized as a key indicator of accounting information reliability and quality [10]. They reflect lapses in internal controls, reporting errors, or managerial opportunism, and are closely linked to overall corporate information quality.

Despite the widespread adoption of digital transformation, its impact on financial restatements and accounting information quality remains underexplored, highlighting the need for systematic investigation. Accordingly, this study adopts a digital technology–enabled information governance perspective and leverages Python-based textual analysis methods to quantify corporate digital transformation from publicly disclosed documents. By combining computational text analysis with financial data, we examine the pathways and mechanisms through which digital transformation influences financial restatements, aiming to reveal how digital technologies enhance information governance and improve the reliability of financial reporting, providing both theoretical insights and practical implications for corporate management and capital market regulation.

The contributions of this study are threefold. First, this study advances research on the microeconomic effects of digital technologies and the digital economy. It demonstrates that digital transformation mitigates earnings manipulation by reducing information asymmetry and agency costs, and by enhancing internal controls, thereby enriching the literature on how digital technologies influence financial management behavior. Second, this study employs Python-based textual analysis to systematically extract and quantify information related to digital transformation from corporate reports, capturing both the depth and breadth of digital transformation activities. By using Python scripts, the approach distinguishes different types of digital initiatives and identifies strategic shifts, thereby providing a more interpretable empirical foundation for examining. Third, this study examines the boundary conditions under which digital transformation exerts its effects. The impact of digital transformation is significantly attenuated in firms with CEO-Chair duality, under high financing constraints, or in SOEs, which partly supports the research of the digital paradox and deepens our understanding of digital governance.

2. Theoretical Analysis and Research Hypotheses

First, digital transformation enhances firms' capabilities in data collection and analysis, effectively mitigating the occurrence of financial restatements from an information management perspective. Based on information economics, digitalization reduces information asymmetry between management and stakeholders by enabling firms to collect data efficiently and accurately through multiple channels, such as online transactions and social media, combined with automated processing technologies [4]. Since financial restatements often arise from data omissions, calculation errors, or inaccurate information processing, digital technologies can significantly

reduce these risks. Big data and AI not only help identify anomalies and potential errors in financial data but also detect patterns and forecast trends, enabling timely correction of issues in financial statements and lowering the risk of restatements.

Second, digital transformation strengthens internal control mechanisms, thereby mitigating financial restatements from a corporate governance perspective. Agency theory suggests that executives, motivated by achieving performance targets or increasing personal compensation, may engage in opportunistic behavior, which leads to financial restatements [11]. Digital transformation enhances the transparency and visualization of internal information, enabling management and regulatory authorities to more effectively detect and monitor executive behaviors. By signaling digital transformation to capital markets and regulatory agencies, firms can also increase the attention of investors, analysts, and government bodies, further constrain managerial opportunism, and thereby decrease the likelihood of financial restatements.

Finally, digital transformation improves operational management and production efficiency, reducing the motivation for earnings restatements. According to the Resource-Based View, digital capabilities are valuable, rare, and hard-to-imitate resources that help firms optimize production processes, enhance efficiency and product quality, and reduce costs [12]. Intelligent management based on data analytics enables firms to better understand customer demands and operational conditions, facilitating more informed managerial decisions [13]. The improvement in operational management not only strengthens the firm's business foundation and financial condition but also increases the incentive to disclose accurate financial information, thereby reducing the likelihood of financial restatements caused by operational pressures or financial distress.

Drawing on the above, this study hypothesizes that:

H1: Digital transformation can reduce the likelihood of financial restatements.

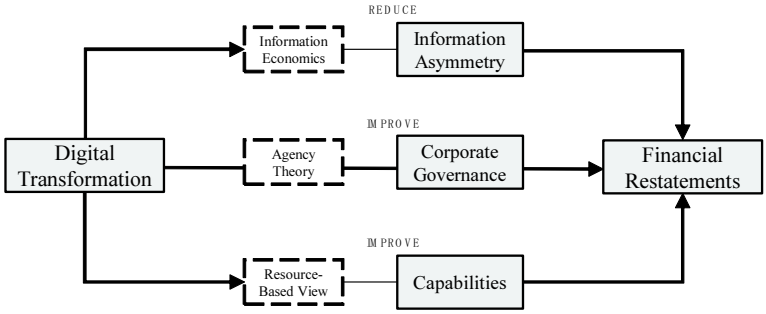


Figure 1. Conceptual Framework.

3. Empirical Design

3.1. Sample Selection and Data Sources

This study examines Chinese A-share listed firms from 2011–2023, applying the following sample filters: (1) financial sector firms are excluded; (2) observations with missing or abnormal data are removed; (3) ST and *ST firms are excluded; and (4) Continuous variables are winsorized at the 1% level to mitigate extreme value effects. After applying these criteria, the final sample comprises 5,186 listed companies, yielding 40,292 observations. The large and diverse sample enhances the reliability and representativeness of the results. Financial and firm-specific data are primarily obtained from the CNRDS and CSMAR databases.

3.2. Variable Measurement

Financial restatement (EIQD) reflects the revision of previously published financial statements due to misstatements, omissions, or errors in reported earnings. The EIQD variable is coded as 1

if the company's annual report for a given year involves a material accounting error that leads to a financial restatement in subsequent years; otherwise, it is coded as 0.

Corporate digital transformation is systematically reflected through the construction of an annual Digital Transformation Index (DIG) for each firm, capturing the depth and breadth of the company's digital strategy. Specifically, following existing literature [14], Python is used to scrape the annual reports of A-share listed companies, evaluating digital transformation from both technological and application perspectives. The focus is on "artificial intelligence technology," "big data technology," "cloud computing technology," "blockchain technology," and "digital technology applications." By identifying keywords and phrases related to digitalization and using the TF-IDF term frequency method, the frequency of these keywords is calculated and log-transformed to derive the composite Digital Transformation Index (DIG).

Table 1. Variable Definitions.

Type	Name	Symbol	Definition
Dependent Variable	Financial Restatement	**EIQD**	Equals 1 if the firm issues a financial restatement in the year; equals 0 otherwise.
Independent Variable	Digital Transformation	**DIG**	Natural logarithm of the total annual frequency of digital-related keywords appearing in the firm's annual reports.
Control Variables	Firm Size	**SIZ**	The natural log of total assets.
	Leverage	**LEV**	Total liabilities scaled by total assets.
	Profitability	**ROA**	Net income over average total assets.
	Firm Age	**AGE**	The natural log of years since establishment.
	Operating Cash Flow	**CFO**	Net cash flow from operations scaled by average total assets.
	Ownership Concentration	**SHA**	The largest shareholder's equity stake.
	Internal Audit	**INAU**	Equals 1 if the firm discloses an internal audit report; 0 otherwise.

To mitigate the influence of potential confounding factors on financial restatements, this study includes eight control variables: firm size (SIZ), leverage (LEV), profitability (ROA), firm age (AGE), operating cash flow (CFO), ownership concentration (SHA), and internal audit disclosure (INAU). Additionally, time and industry dummies are incorporated to absorb unobserved heterogeneity across years and sectors.

4. Empirical Results and Analysis

4.1. Descriptive Statistics

As shown in Table 2, the dataset contains 40,292 firm-year observations. The mean value of financial restatement (EIQD) is 0.288, suggesting that 28.8% of sample firms restated their financial reports. The mean value of digital transformation (DIG) is 2.961, with a median of 2.944, suggesting no significant skewness. Regarding the control variables, the average firm size is 22.23; the mean leverage ratio is 44.7%. The average return on assets is 2.9%, indicating that the profitability of listed firms is relatively low. The mean value of AGE is 2.041, implying an average firm age of approximately eight years. Operating cash flow accounts for 4.2% of total assets on average. The largest shareholder holds, on average, 33.65% of shares. Additionally, 97.8% disclose internal audit reports.

Table 2. Descriptive Statistics.

	N	Mean	Median	Standard Deviation
EIQD	40292	0.288	0	0.453
DIG	40292	2.961	2.944	1.271
SIZ	40292	22.23	21.99	1.512
LEV	40292	0.447	0.417	1.017
ROA	40292	0.029	0.036	0.677
AGE	40292	2.041	2.197	0.957
CFO	40292	0.042	0.043	0.102
SHA	40292	33.65	31.14	15.03
INAU	40292	0.978	1	0.147

4.2. Regression Analysis

Table 3 reports the baseline regression results. As shown in Columns (1) – (3), the estimated effects are stable and robust across alternative model specifications. Column (1) includes digital transformation (DIG) as the sole explanatory variable. Column (2) adds firm-level control variables, while Column (3) further incorporates year and industry fixed effects. Across all specifications, the coefficient on DIG remains negative and statistically significant. In Column (3), DIG is significantly negative ($\beta = -0.023$, $p < 0.01$), indicating that a one-unit increase in digital transformation reduces the probability of financial restatements by approximately 2.3 percentage points on average, accounting for 8% of the mean value of financial restatements. Overall, the results in Table 3 suggest that corporate digital transformation significantly reduces the likelihood of financial restatements, supporting H1.

Table 3. Regression Results.

	(1)	(2)	(3)
VARIABLES	EIQD	EIQD	EIQD
DIG	-0.035***	-0.016***	-0.023***
	(-15.981)	(-7.255)	(-8.705)
SIZ		0.026***	0.021***
		(10.895)	(7.672)
LEV		0.007***	0.007***
		(3.594)	(3.365)
ROA		0.037***	0.037***
		(11.226)	(11.105)
AGE		-0.152***	-0.160***
		(-46.755)	(-45.341)
CFO		0.499***	0.488***
		(22.723)	(22.083)
SHA		0.006***	0.006***
		(26.393)	(26.567)
INAU		0.066***	0.076***
		(4.752)	(5.435)
Year	NO	NO	YES
Industry	NO	NO	YES
Constant	0.446***	-0.212***	-0.049
	(54.784)	(-4.131)	(-0.750)
Observations	40,292	40,292	40,292
Adjust R^2	0.0117	0.0734	0.0771

Note： *, **, *** denote 10%, 5%, 1% significance; t-statistics in parentheses. Same below.

4.3. Robustness Tests

Alternative Measures of the Independent Variable. This study constructs two alternative proxies for digital transformation. First, based on intangible asset notes, items containing keywords such as "software," "management system," "intelligent platform," "data," and "artificial intelligence" are aggregated at the firm-year level and divided by total intangible assets, yielding a resource-investment-based measure. Second, from the perspective of external information intermediaries, an analyst-attention-based measure is constructed by counting annual analyst reports mentioning keywords like "digitalization," "big data," and "artificial intelligence," with the count log-transformed after adding one. Regression results in Columns (1) and (2) of Table 4 show that both measures produce results consistent with the main regression, confirming robustness.

Propensity Score Matching (PSM) Method. To address potential unobservable factors, the study applies PSM. Firms are divided into a treatment group (undergoing digital transformation) and a control group (not transformed). Propensity scores are estimated via a Logit model using control variables, followed by one-to-one nearest neighbor matching. Post-matching, no significant differences remain across covariates. Regressions on the matched sample (Column 3, Table 4), confirming the robustness of digital transformation's effect on financial restatements.

Instrumental Variable Approach. Given the potential endogeneity issues, this study uses regional digital economy development as an instrumental variable (IV) and estimates the model using two-stage least squares (2SLS). Regional digital economy development is measured using four indicators: broadband penetration, proportion of workers in the tech industry, per capita telecommunication volume, and mobile phone users per 100 people. These indicators are standardized using principal component analysis to form instrumental variable (IV). This index is highly correlated with the degree of corporate digital transformation but does not directly affect financial restatements, satisfying the relevance and exogeneity requirements for an instrumental variable. The 2SLS estimates in Columns (4) – (5) confirm that the negative DIG effect persists after controlling for endogeneity, reinforcing the robustness of our findings.

Table 4. Regression Results of Robustness Tests.

	(1)	(2)	(3)	(4)	(5)
VARIABLES	EIQD	EIQD	EIQD	DIG	EIQD
DIG	-0.001***	-0.001***	-0.023**		-0.112***
	(-3.451)	(-2.878)	(8.610)		(-3.594)
IV				0.021***	
				(3.821)	
Controls	YES	YES	YES	YES	YES
Year	YES	YES	YES	YES	YES
Industry	YES	YES	YES	YES	YES
Constant	0.063***	0.135***	1.044***	-2.631***	-0.529***
	(6.799)	(13.477)	(15.861)	(-17.846)	(-7.165)
Observations	40,292	40,292	39,946	40,292	40,292
Adjust R^2	0.225	0.174	0.185	0.415	0.156

4.4. Mechanism Analysis

To systematically examine the underlying mechanisms, this study constructs a mediation model to empirically quantify the impact of digital transformation on financial restatements. First, digital transformation improves information processing and transmission efficiency, narrowing the information gap between management and external investors, thereby alleviating information asymmetry. ASY is extracted as the first principal component of liquidity ratio, illiquidity ratio, and return reversal. Specifically, the liquidity ratio and illiquidity ratio reflect the liquidity and stability of a company's assets, while the return reversal measure reflects changes in market reactions. The combination of these three variables helps to reveal the transparency and liquidity

of information, thereby measuring the information gap between management and external investors. Columns (1) and (2) of Table 5 show that, through this mediation channel, the indirect effect of digital transformation on financial restatements is −0.000495, explaining 2.06% of the total effect. This result indicates that digital transformation mitigates financial restatement risk by reducing information asymmetry.

Second, digital transformation enhances data processing capabilities and information-sharing channels, strengthening external oversight and reducing agency costs. In this study, asset turnover is used to measure agency costs (AGC), calculated as operating revenue divided by average total assets. Asset turnover reflects the efficiency of a firm's resource utilization and indirectly reveals managerial efficiency and incentive structures in operational decision-making. Columns (3) and (4) of Table 5 indicate that, through the mediation of agency costs, the indirect effect of digital transformation on financial restatements is −0.003726, explaining 14.9% of the total effect. This suggests that digital transformation can also reduce the risk of financial restatements by alleviating agency problems, thereby enhancing the overall effectiveness of corporate information governance.

Third, digital transformation strengthens internal controls by improving the information environment and technology, thereby reducing financial restatements. This study introduces the Internal Control Quality (ICQ) variable to measure the level of internal controls. Specifically, ICQ is constructed based on whether a firm discloses internal control deficiencies; a value of 1 is assigned if no deficiencies are disclosed, and 0 if deficiencies are disclosed (data sourced from the CNRDS database). The regression results in Columns (5) and (6) of Table 5 show that through the mediation of internal control quality, the indirect effect of digital transformation on financial restatements is 0.001027, explaining 4.67% of the total effect. So digital transformation significantly improves internal control quality, which, in turn, enhances the authenticity and reliability of financial reporting.

Table 5. The Mediating Effect Test of the Influence Mechanism.

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	ASY	EIQD	AGC	EIQD	ICQ	EIQD
DIG	-0.011***	-0.024***	-0.069***	-0.025***	0.013***	-0.022***
	(-5.160)	(-3.852)	(-4.681)	(-4.061)	(4.954)	(-8.255)
ASY		0.045*				
		(1.931)				
AGC				0.054***		
				(4.227)		
ICQ						-0.079***
						(-5.239)
Controls	YES	YES	YES	YES	YES	YES
Year	YES	YES	YES	YES	YES	YES
Industry	YES	YES	YES	YES	YES	YES
Constant	0.007**	0.013***	0.011***	0.019***	1.074***	0.039
	(2.162)	(4.268)	(3.362)	(6.491)	(15.498)	(0.598)
Observations	40,292	40,292	40,292	40,292	40,292	40,292
Adjust R^2	0.492	0.0776	0.509	0.0778	0.0460	0.0825

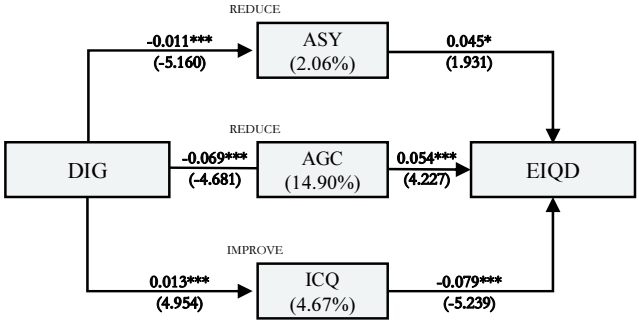


Figure 2. Mechanism Diagram.

4.5. Heterogeneity Analysis

Although digital transformation is widely viewed as a tool for improving information processing and corporate governance, prior studies highlight a "digital paradox," whereby digital investments do not necessarily translate into improved governance or firm performance. This paradox arises because digital technologies are essentially neutral tools, and their governance effects depend critically on how firms deploy them and on the institutional and organizational environments in which they operate. Accordingly, this study further conducts heterogeneity analyses to test if firm environments constrain the impact of digital transformation.

First, when the roles of board chair and CEO are combined, internal power concentration weakens board oversight, limiting the extent to which enhanced information transparency from digital technologies can translate into effective governance. Based on whether firms exhibit CEO duality (DUAL), this study conducts subsample regressions. Table 6, Columns (1) – (2), show that digital transformation significantly reduces the likelihood of financial restatements only in firms without CEO duality, whereas the suppressing effect is not significant in firms with CEO duality. These findings indicate that when internal governance structures are weakened, the governance potential of digital technologies may be substantially undermined.

Second, financing constraints limit firms' sustained investments in information systems and digital infrastructure, making it difficult for digital transformation to develop into systematic and standardized information-processing mechanisms. Financing constraints are proxied by the SA index, which is constructed based on firm size (Size) and firm age (Age), with larger values indicating more severe financing constraints. The sample is split at the median SA index into high- and low-constraint subsamples. The results (Table 6) show that digital transformation significantly reduces financial restatements only among firms with low financing constraints. These findings suggest that a favorable financing environment is essential for digital transformation to improve corporate governance and financial reporting quality.

Firms with different ownership structures differ in financing channels, operational objectives, and regulatory environments. SOEs are politically sensitive and subject to scrutiny by the government, media, and the public. Their performance often serves a political or reputational function, but multiple objectives and the separation of ownership and control can lead to internal governance issues. Results in Columns (5) – (6) of Table 6 indicate digital transformation significantly reduces financial restatements in both SOEs and non-SOEs, with a substantially larger coefficient for non-SOEs. This may be because SOEs' special ownership structure and internal control constraints limit the effectiveness of digital transformation in improving transparency and governance, whereas digital transformation can more fully realize its governance function in non-SOEs, resulting in a stronger mitigation of financial restatements.

Table 6. Regression Results of Heterogeneity Analysis.

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	DUAL	non-DUAL	High-SA	Low-SA	SOEs	non-SOEs
DIG	-0.004	-0.029***	0.005	-0.043***	-0.017***	-0.025***
	(-0.828)	(-9.586)	(1.274)	(-12.408)	(-3.826)	(-7.841)
Controls	YES	YES	YES	YES	YES	YES
Year	YES	YES	YES	YES	YES	YES
Industry	YES	YES	YES	YES	YES	YES
Constant	-0.340***	-0.047	-0.081	-0.267***	1.630***	0.973***
	(-2.674)	(-0.641)	(-0.987)	(-2.786)	(14.908)	(11.882)
Observations	11,345	28,947	20,146	20,146	13,624	26,668
Adjust R^2	0.100	0.068	0.051	0.075	0.033	0.102

5. Conclusion

Digital transformation has become a prominent topic in contemporary economic development [13], attracting considerable attention for its impact on corporate financial management and information governance. This study examines how digital transformation affects corporate financial reporting quality, using Chinese A-share listed firms from 2011–2023, combining theoretical analysis with empirical testing. The findings demonstrate that digital transformation significantly reduces the occurrence of financial restatements, with results remaining robust across multiple checks. Mechanism analysis reveals that digital transformation improves governance through mitigating information asymmetry, curbing agency costs, and optimizing internal controls. Furthermore, the impact is weaker among firms with CEO-chair duality, financing constraints, and state ownership, highlighting the important moderating roles of firm environment and characteristics in the effectiveness of digital transformation.

As an emerging managerial practice, digital transformation poses new challenges to traditional financial management theories. Through theoretical analysis and empirical testing, this study demonstrates that digital transformation can effectively mitigate financial restatements, thereby enrich the financial management literature and establish a comprehensive analytical framework that integrates financial information acquisition with internal control optimization. Furthermore, heterogeneous effects across firm types reveal how firm characteristics interact with digital transformation. Digital transformation substantially elevates financial reporting quality, reduces the risk of financial restatements, and improves transparency and reliability, thereby strengthening investor confidence and decision-making value. By leveraging digital systems, firms can access financial information more quickly and accurately, while optimizing internal controls to minimize financial misstatement and operational risks. These findings offer practical guidance for improving financial management efficiency, enhancing investor trust, and attracting investment capital.

References

[1] Millán JM, Lyalkov S, Burke A, Millán A, van Stel A. 'Digital divide' among European entrepreneurs: Which types benefit most from ICT implementation? J Bus Res. 2021;533-547.

[2] Babina T, Fedyk A, He A, Hodson J. Artificial intelligence, firm growth, and product innovation. J Financ Econ. 2024;103745.

[3] Jafari-Sadeghi V, Garcia-Perez A, Candelo E, Couturier J. Exploring the impact of digital transformation on technology entrepreneurship and technological market expansion: The role of technology readiness, exploration, and exploitation. J Bus Res. 2021;100-111.

[4] Gebauer H, Fleisch E, Lamprecht C, Wortmann F. Growth paths for overcoming the digitalization paradox. Bus Horiz. 2020;313-323.

[5] Song D, Tan Z, Wang W, Zhai RX. Digital transformation and corporate social responsibility engagement: Evidence from China. Int Rev Financ Anal. 2025;103805.
[6] Lateef A, Omotayo FO. Information audit as an important tool in organizational management: A review of literature. Bus Inform Rev. 2019;15-22.
[7] Huang H, Wang C, Wang L, Yarovaya L. Corporate digital transformation and idiosyncratic risk: Based on a corporate governance perspective. Emerg Mark Rev. 2023;101045.
[8] Hu J, Li AY, Zhang FF. Does accounting conservatism improve the corporate information environment? J Int Account Audit Tax. 2014;32-43.
[9] Saleh WA, Mansour M. Is audit committee busyness associated with earnings management? The moderating role of foreign ownership. Account Res J. 2024;80-97.
[10] Khuong NV, Liem NT, Anh LHT, Dung BTN. Related party transactions, accrual-based earnings management, and real activities earnings management in emerging markets. Pac Account Rev. 2024;60-76.
[11] Zhang L, Shan YG, Chang M. Can CSR disclosure protect firm reputation during financial restatements? J Bus Ethics. 2021;157-184.
[12] Chen D, Esperança JP, Wang SF. The impact of artificial intelligence on firm performance: An application of the resource-based view to e-commerce firms. Front Psychol. 2022;884830.
[13] Tsakalerou M. Tailoring digital transformation: A customized DESI framework for economic and societal growth. Telematics Inform Rep. 2025;100244.
[14] Jiang K, Du X, Chen Z. Firms' digitalization and stock price crash risk. Int Rev Financ Anal. 2022;102196.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)

© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261126

Leap Forward or Proceed Step by Step? Regional Development Effects and Path Differentiation of China's New Quality Productivity

Yan-juan Hu[a,*]

[a] *School of Materials Science and Engineering, South China University of Technology, China*

Abstract: Addressing the strategic task of high-quality development, new quality productive forces have become the core driving force for regional economic transformation and upgrading. However, how to scientifically evaluate the formation process and effectiveness of new quality productive forces has become a critical and urgent challenge. Based on the conceptual characteristics and generative logic of new quality productive forces, this study constructs an evaluation framework encompassing dimensions such as scientific and technological innovation, industrial transformation, production relations, and factor allocation. The Delphi method and the Analytic Hierarchy Process (AHP) were employed to determine the weights of indicators at each level. Taking 31 provincial-level administrative regions in China as research samples, the study draws on multi-source statistical data to conduct an empirical measurement of the development level of new quality productive forces and to classify regional development types. The findings reveal: The findings reveal pronounced regional gradient differences, forming four echelon tiers (leading, active, steady, and emerging), with significant heterogeneity in development paths and dominant driving factors across regions. Moreover, the coupling coordination between driving pathways and development outcomes is generally low nationwide, with marked disparities in input-output conversion efficiency. This research helps deepen the understanding of the formation mechanisms and developmental performance of NQPF from a full-chain perspective, providing empirical evidence for promoting the development of NQPF tailored to local conditions and optimizing regional high-quality development policies.

Keywords: New quality productive forces; driving pathway; development outcome; evaluation index system

1. Research Question and Literature Review

Against the backdrop of the deep evolution of a new round of technological revolution and industrial transformation, and the shift of China's economic development stage from

* Corresponding Author, Yanjuan Hu, South China University of Technology School of Materials Science and Engineering, China; Email: huyj@ibmd.org.cn.

Funded Project: Major Research Project of the Guangdong Institute of China Engineering Science and Technology Development Strategy: Research on the Leapfrog Development Strategy and Evaluation System of New Quality Productivity in Guangdong Province (Project No.: 2025-DFZD-25)

high-speed growth to high-quality development, developing NQPF serves as a key lever for shaping new developmental advantages, cultivating and expanding new drivers of growth, and accelerating the construction of a modern industrial system. China's 15th Five-Year Plan emphasizes "developing NQPF tailored to local conditions." As various regions accelerate their layout around the digital economy, strategic emerging industries, future industries, and green transformation, a more realistic and urgent issue has come to the forefront: what level of NQPF development should ultimately be achieved, what primarily drives its formation, and whether it can be truly translated into sustainable efficiency gains and improvements in people's livelihoods? Given the current rise in macroeconomic uncertainty and tightening local fiscal constraints, it is even more necessary to strengthen objective measurement to "account for inputs and outcomes," enabling the precise allocation of limited resources to bottleneck areas and high-efficiency sectors, while preventing a "reckless rush," redundant construction, and inefficient investment. A quantifiable, comparable, and traceable evaluation system for NQPF is both a theoretical foundation for answering the above questions and a key instrument for policy implementation.

As a large country with vast territory and significant regional disparities, how effective is the development of NQPF across different regions of China? What characteristics does it exhibit? What driving pathways have been adopted? These issues warrant further evaluation and interpretation. New economic geography provides an important perspective for understanding the spatial distribution of new quality productive forces and their regional disparities. Relevant literature primarily focuses on dimensions such as spatial agglomeration, factor mobility, spatial spillovers, and regional imbalances [1]. With the development of the digital economy, the cross-regional flow of new factors such as data and talent has weakened the constraints of geographical distance. However, the agglomeration effect of innovation factors continues to exacerbate the "Matthew effect" across regions [2]. The "core-periphery" model indicates that the mismatch between population and industrial agglomeration across regions, spatial linkages, and regional specialization are important micro-level drivers that widen regional disparities in new quality productive forces [3]. Digital technologies can transcend the constraints of geographical distance and, through technological diffusion and the sharing enabled by digital platforms, rapidly replicate and extend advanced intelligent technologies to different regions. However, regional disparities persist in terms of improving factor productivity [4]. In the United States, the areas with the highest concentration of robot employment are the traditional manufacturing hubs in the Midwest, where relevant industrial clusters and other related industries are geographically proximate to one another, exhibiting a regionalized and specialized innovation economy pattern [5].

Against the backdrop of China, existing research has measured NQPF mainly from the dimensions of productive factor structure, efficiency performance, industrial forms, and development concept characteristics, forming several representative measurement paradigms. The first is an improved measurement paradigm based on the three traditional factors of production [6]. The second understands NQPF as the comprehensive outcome of the accelerated development of new technologies, new economies, and new business formats, characterizing NQPF levels through structural changes at the industrial level. Shi and Yu constructed an NQPF measurement indicator system from three aspects: new manufacturing, new services, and new business formats [7]. The third directly attributes the core connotation of NQPF to the continuous improvement of TFP, thereby using TFP or its extended forms as a proxy variable for NQPF [8]. Building on traditional TFP measurement, environmental constraints are introduced, and GTFP is used to represent

the development level of NQPF [9]. The fourth conducts a multi-dimensional comprehensive measurement of NQPF based on developmental goals and value orientations [10].

However, existing research on productivity evaluation often tends to adopt a single perspective focused on either the input side or the output side within its analytical framework [11]. A comprehensive measurement framework that organically integrates driving pathways and development outcomes is still lacking. This makes it difficult to systematically examine the generation mechanism and value destination of the "input-transformation-outcome" process. Therefore, this paper innovatively proposes an integrated measurement framework of "driving pathways–development outcomes," further disaggregated into 5 primary indicators, 13 secondary indicators, and 42 tertiary indicators. Based on this framework, multi-source data are used to comprehensively evaluate the actual level, spatial distribution, and gradient pattern of NQPF across 31 provincial-level regions, to examine the degree of matching between driving pathways and development outcomes, and subsequently to identify areas of strength and weakness, providing a benchmark for cross-regional experience sharing and healthy competition. This study involves determining the weights of multidimensional indicators, which requires integrating the knowledge and experience of experts from multiple fields to reflect the relative importance among indicators. The Delphi method can effectively consolidate expert consensus, while the AHP can transform qualitative judgments into quantitative weights and conduct consistency tests. This study provides a more integrated measurement framework, which helps to analyze the driving pathways and developmental outcomes of new quality productive forces in the Chinese context. From the perspectives of regional agglomeration and spatial spillover, it reveals the gradient differences and type heterogeneity in the development of regional new quality productive forces, offering empirical insights for developing new quality productive forces.

2. Theoretical Framework and Research Methods

NQPF are generated by revolutionary technological breakthroughs, innovative allocation of production factors, and deep industrial transformation and upgrading. Their basic connotation is the leapfrog development of laborers, means of labor, objects of labor, and their optimized combination. Their core symbol is a significant increase in TFP, characterized by innovation and high quality, and their essence is advanced productivity. This paper defines the development process of NQPF as a systemic process that moves from input through transformation to the ultimate release of value. NQPF are viewed as the comprehensive result generated by driving pathways such as factor input, technological innovation, and industrial transformation, and guaranteed by specific production relations and institutional environments, ultimately manifested as improvements in economic efficiency, structural optimization, and social welfare. The development level of NQPF is characterized from both the generation mechanism and actual performance dimensions, as illustrated in Figure 1.

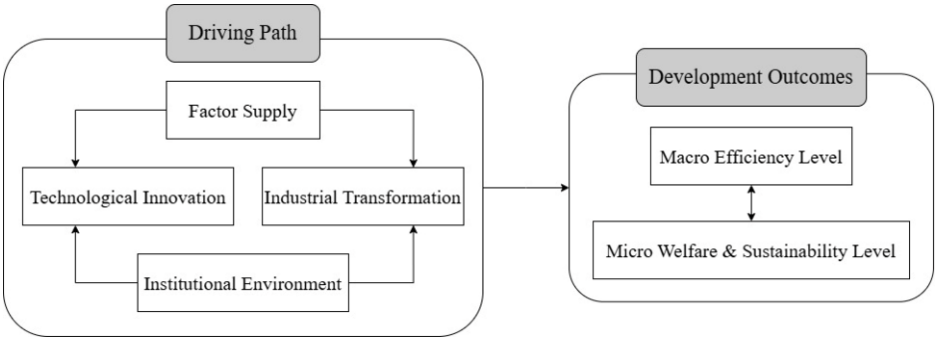


Figure 1. Theoretical Model for Measuring NQPF

Based on the inherent logic of "driving pathways–development outcomes," an evaluation framework is constructed encompassing five primary dimensions: factor supply, technological innovation, industrial transformation, production relations, and development outcomes. This framework is further disaggregated into 13 secondary indicators and 42 tertiary indicators. To ensure the scientific validity and authority of the indicator weights, this study adopted a combined Delphi and Analytic Hierarchy Process (AHP) approach to determine the weightings. An evaluation panel was formed, comprising 17 experts from institutions including the Institute of Market Economy of the Development Research Center of the State Council, the Guangdong Research Institute of China's Engineering Science and Technology Development Strategy, the Department of Industry and Information Technology of Guangdong Province, Tsinghua University, the University of Electronic Science and Technology of China, and South China University of Technology. The expert selection criteria were as follows: (1) at least 10 years of professional experience in relevant fields; (2) holding an associate senior professional title or above, or an equivalent level of expertise; and (3) direct involvement in industrial policy formulation, science and technology innovation management, or theoretical research in industrial economics. A two-round Delphi consultation process was implemented: in the first round, an open-ended questionnaire was distributed to collect experts' independent judgments and supplementary comments on the importance of each indicator; in the second round, the statistical results and anonymized summaries of opinions from the first round were provided, allowing experts to reconsider and revise their ratings. Both rounds achieved a 100% response rate. A pairwise comparison matrix was constructed using the AHP, and the random consistency ratio (CR) for all matrices remained below 0.10, thereby passing the consistency test. The final weights were determined by calculating the arithmetic mean of the valid ratings from the second round. The coefficient of variation for each indicator's weight ranged from 0.07 to 0.15, indicating a satisfactory degree of expert opinion convergence and a high level of result reliability.

Regarding the selection of measurement units, this paper takes provincial-level administrative regions as the basic unit of analysis. Based on practical considerations of data completeness and availability, the study covers 31 provincial-level administrative regions in mainland China, including provinces, autonomous regions, and municipalities directly under the central government, excluding the Hong Kong Special Administrative

Region, the Macau Special Administrative Region, and the Taiwan region.[†] A multi-source data fusion approach is adopted to construct the indicator database, relying on official statistics as the foundation, authoritative releases from ministries and commissions as support, and third-party rankings and big data as supplements, thereby enhancing the authority, completeness, and practical sensitivity of the indicator measurement. The indicators related to basic data are primarily sourced from the China Statistical Yearbook, the China Statistical Yearbook on Science and Technology, and the statistical bulletins on national economic and social development of each province, which are used to obtain key variables such as gross regional product, R&D expenditure. Some indicators are directly obtained from the official websites and public data platforms of government departments. Furthermore, to address the limitations of traditional statistical data in terms of timeliness and micro-level characterization, this study incorporates enterprise big data and web-based data collection methods. Subsequently, the Min-Max Normalization method is applied to standardize all indicators dimensionlessly. Combined with the indicator weights determined above, the comprehensive NQPF scores and sub-item scores for each provincial-level administrative region are calculated using the weighted summation method, as shown in Table 1.

Table 1 Evaluation Indicator System and Data Sources for New Quality Productive Forces Development

Primary indicators	Secondary indicators	Third-level indicators	Data sources and calculations
Technological innovation	Strategic technological strength	National laboratories	Network Collection
		A high-level research university	Network Collection
		National R&D institutions	Network Collection
		A leading technology enterprise	Web crawler
	Innovation investment	R&D expenditure as a proportion of regional GDP	Statistical Yearbook and Statistical Bulletin
		Proportion of basic research funding to R&D expenses	Statistical Yearbook and Statistical Bulletin
		Proportion of enterprise R&D expenses to R&D expenses	Statistical Yearbook and Statistical Bulletin
	Innovative output	Invention patents per 10,000 people	Derwent and National Intellectual Property
		Number of National Science and Technology Awards received	Based on the latest year's award statistics
		Proportion of future industry patents in invention patent applications	Derwent and National Intellectual Property
		Number of AI large model filings	Data from the Cyberspace Administration of China
		Technology contract transaction amount	Statistical Yearbook and Statistical Bulletin
Industrial transformation	Enterprise clustering	Number of unicorn companies worldwide	Hurun Research Institute
		Number of national-level "Little Giant" specialized and innovative enterprises	Cumulative data up to the time of collection, Ministry of Industry and Information Technology website
		Number of high-tech enterprises	Cumulative data up to the time of collection
	Structural optimization	The proportion of the digital economy scale in regional GDP	Statistical Yearbook, Statistical Bulletin, China Academy of Information and Communications Technology

[†] According to the latest data from the National Bureau of Statistics, the 31 provinces, autonomous regions, and municipalities directly under the central government (excluding Hong Kong, Macao, and Taiwan) covered in this study account for approximately 99.5% or more of China's total population. Specifically, as of the sample period (2020–2023), the total population of the 31 administrative regions represents about 99.6% of the national population, essentially covering the vast majority of China's population and demonstrating good representativeness.

Primary indicators	Secondary indicators	Third-level indicators	Data sources and calculations
		The proportion of enterprises in emerging industries among national-level "Little Giant" specialized and innovative enterprises	Calculated based on collected data
		The proportion of the added value of the producer service industry in the added value of the tertiary industry	Statistical Yearbook and Statistical Bulletin
	Green development	Number of national-level green industrial parks	Cumulative data up to the time of collection, Ministry of Industry and Information Technology website
		Number of national-level green factories	Cumulative data up to the time of collection, Ministry of Industry and Information Technology website
		Number of national-level green supply chain management enterprises	Cumulative data up to the time of collection, Ministry of Industry and Information Technology website
Supply of factors	Capital elements	Scale of government guidance funds	Sina Finance
		Market-oriented fund of funds scale	Sina Finance
		Number of banking financial institutions	National Financial Regulatory Administration
		Digital Financial Inclusion Index	Data from the Digital Finance Research Center of Peking University
	Talent elements	Proportion of the population with a bachelor's degree or above in the total population	Statistical Yearbook
		R&D personnel are full-time equivalent	Statistical Yearbook
		Top 2% of global scientists	Latest Stanford University rankings
		Number of highly cited scientists worldwide	Latest data from Clarivate
	Data elements	Scale of the intelligent computing center	Search online for the scale of computing power centers across different regions
		Scale of government data openness	Local government data open platforms
		Number of big data (potential) unicorn companies	China Big Data Unicorn Enterprise List
		Establishment of data trading institutions	According to data from the Telecommunications Industry Report, the number of data trading institutions set up is 1, and the number of none, is 0.
Production relations	Business environment	Market entity entry rate	Number of newly established enterprises / Total number of enterprises, Qichacha Cha
		Government digital service capabilities	Report on the Development of Digital Service Capabilities of Local Governments in China
	Open to the outside world	The proportion of actual utilization of foreign capital to regional GDP	Statistical Yearbook and Statistical Bulletin
		Number of foreign R&D centers	Online search
Development Achievements	Industry efficiency	Overall labor productivity	Replace it with per capita GDP, statistical yearbooks and statistical bulletins
		Total factor productivity	The average total factor productivity of listed companies in each region
	The people are happy	Urban Registered Unemployment Rate	Statistical yearbook, inverse indicators
		Growth rate of per capita disposable income of residents	Statistical yearbook, nominal growth rate
		Growth rate of per capita consumption expenditure by residents	Statistical yearbook, nominal growth rate

3. Measurement Results and Regional Differences in NQPF Development

Hierarchical clustering results indicate that the development of NQPF in China exhibits significant echelon differentiation characteristics, forming an overall pattern of geese-formation advancement. Provincial-level administrative regions can be categorized into four types: leading development, active development, steady development, and emerging development. The overall pattern is "5+5+8+13," showing an expanding head advantage, a contracting middle echelon, and an overall evolution from an olive-shaped to a multipolar structure, as shown in Figure 2.

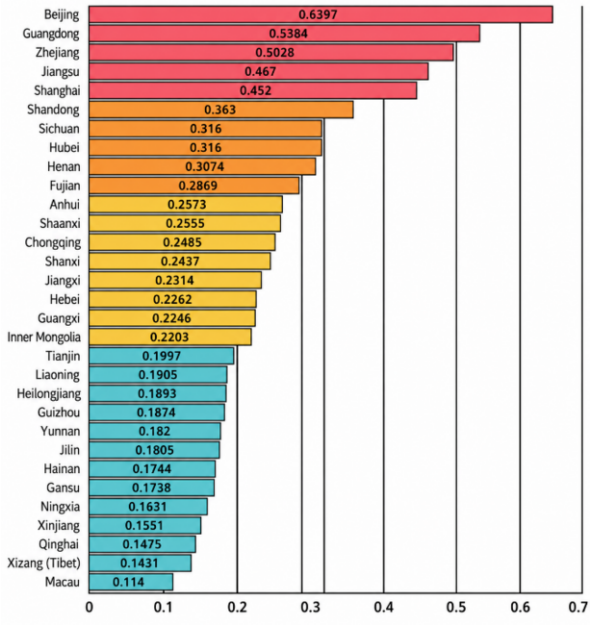


Figure 2. Ranking of Total NQPF Development Scores by Provincial Administrative Region, 2025

The leading development echelon comprises five provincial-level administrative regions, including Beijing, Guangdong Province, Jiangsu Province, Shanghai, and Zhejiang Province, with comprehensive scores ranging from 0.4520 to 0.6397. Beijing maintains the top position nationwide with a score of 0.6397, but its gap with Guangdong Province (0.534) has significantly narrowed. These two provinces lead the other regions considerably in comprehensive scores, constituting the core leading force of the first echelon.

The active development echelon comprises Shandong Province, Fujian Province, Hubei Province, Tianjin, and Anhui Province, with comprehensive scores ranging from 0.2869 to 0.3630. This echelon has overall surpassed the medium development level, with NQPF beginning to form a large-scale foundation and making relatively rapid progress in the transformation and upgrading of traditional industries and the cultivation of emerging industries. However, compared with the leading echelon, significant gaps remain in original innovation capacity, high-end factor agglomeration, and full-chain coordination, placing this echelon overall in an accelerated catch-up phase.

The steady development echelon includes Chongqing, Hunan Province, Sichuan Province, Shaanxi Province, Jiangxi Province, Hebei Province, Henan Province, and Inner Mongolia Autonomous Region, with comprehensive scores concentrated between 0.2203 and 0.2573. This echelon covers the central region and some key western provinces, serving as a critical carrying area for the transition of NQPF from point-based breakthroughs to area-wide expansion. These regions have relatively solid industrial foundations but still face shortcomings in the supply of high-end innovation resources and the efficiency of technology conversion.

The emerging development echelon comprises 13 provincial-level administrative regions, mainly distributed across the northeast, northwest, and southwest regions, with comprehensive scores generally below 0.20. Affected by geographical conditions, economic development stages, and long-term structural constraints, these regions generally face challenges such as the outflow of high-end talent, insufficient supply of innovation capital, and an industrial structure heavily reliant on resource-based and traditional industries.

In summary, the numbers of provinces and their regional distributions under the above different classifications are shown in Table 2. It is worth noting that the highest score in the active development echelon is 0.3630, while the lowest score in the leading development echelon is 0.4520, creating a clear score gap of nearly 0.09. This indicates the difficulty of achieving cross-echelon leaps in NQPF, highlighting the persistent prominence of interregional developmental imbalances.

Table 2 the classification of the 31 analyzed administrative regions

Category	Number of Regions	Percentage (%)
Leading Development	5 (Beijing, Guangdong, Jiangsu, Shanghai, Zhejiang)	16.1
Active Development	5 (Shandong, Fujian, Hubei, Tianjin, Anhui)	16.1
Stable Development	8 (Chongqing, Hunan, Sichuan, Shaanxi, Jiangxi, Hebei, Henan, Inner Mongolia)	25.8
Emerging Development	13 (Liaoning, Xinjiang, Hainan, Guizhou, Guangxi, Shanxi, Jilin, Yunnan, Heilongjiang, Ningxia, Qinghai, Gansu, Tibet)	41.9
Total	31	100.0

4. Internal Elements and Driving Pathway Differences in NQPF

In the factor supply dimension, regional differentiation is most pronounced, with resource agglomeration effects highly concentrated, presenting a clear bipolar pattern. Beijing ranks first nationwide with a score of 0.1794, forming a prominent leading advantage, followed by Shanghai with a score of 0.1220. These two cities are highly agglomerated in terms of high-end human capital, research resources, and financial capital allocation. On one hand, they host dense clusters of top-tier universities and research institutions, continuously supplying high-level innovative talent; on the other hand, as the national financial management center and an international financial center, respectively, they hold pivotal positions in venture capital, patient capital, and multi-tiered capital market system development.

In the technological innovation dimension, Beijing maintains its national lead with a score of 0.2352, an advantage significantly higher than provinces in the second echelon, highlighting its systemic advantages in basic research, original innovation, and the deployment of national strategic scientific and technological forces, making it the core innovation source for NQPF nationwide.

In the industrial transformation dimension, overall performance is the most prominent, serving as the main driving force for current NQPF development. Guangdong Province ranks first nationwide with a score of 0.2017, surpassing Beijing (0.1284) to become the leader in this dimension. Its strong manufacturing base, complete industrial chain system, and advantages in industrial clusters such as next-generation information technology and high-end equipment provide it with significant advantages in the conversion and large-scale application of scientific and technological achievements.

In the production relations dimension, overall scores are relatively low, but regional differences are comparatively small, with the best balance among the five dimensions. Hainan Province ranks first nationwide with a score of 0.0488, followed closely by Shanghai. Hainan's advantage primarily stems from its pioneering explorations in market access, tax arrangements, and institutional rule alignment through the construction of the free trade port, providing a differentiated institutional environment for NQPF development. Overall, production relations remain a key constraint on NQPF development, yet they also represent the area with the greatest potential for improvement.

In the development outcomes dimension, regional rankings exhibit characteristics significantly different from other dimensions. Three provinces enter the "leading development" echelon in this dimension, with western regions such as Xinjiang (0.0692) and Xizang (0.0676) scoring high, even surpassing some developed eastern provinces. Driven by a combination of national policy support and low-base effects, western regions demonstrate strong catch-up momentum in labor productivity growth, resident income growth, and livelihood improvement, reflecting the phased achievements of NQPF diffusion to less developed areas.

5. Development Types of NQPF "Input-Output" Transformation

The development of NQPF across provincial-level regions in China exhibits significant uneven characteristics. The correlation coefficient between the driving pathway score (aggregated from the four dimensions of factor supply, technological innovation, industrial transformation, and production relations) and the development outcome score (reflecting final development results) is only 0.1616, indicating extremely low coupling between the two. This suggests that high-intensity inputs in capital, talent, technology, and institutions across regions do not necessarily translate into corresponding improvements in economic efficiency and social welfare, revealing a widespread "input-output" transformation efficiency problem in NQPF development.

The first type is the "high input, high output" efficient leading type, represented by Beijing, Jiangsu Province, and Zhejiang Province. These regions excel in both driving pathway and development outcome dimensions. The second type is the "high input, low output" transformation bottleneck type, represented by Guangdong Province, Hubei Province, and Sichuan Province. These regions rank among the top nationally in driving pathway scores but have significantly lower development outcome scores, revealing a pronounced divergence between driving pathways and development outcomes that collectively reflects the transformation efficiency problem in NQPF development. The

third type is the "low input, high output" catch-up and surpass type, mainly concentrated in the central-western and border regions, such as Xizang, Xinjiang, Yunnan, and Shanxi. These regions have relatively low driving pathway scores but relatively high development outcome scores, presenting a "low input, high output" characteristic. Benefiting from national regional coordinated development strategies, paired assistance policies, and low development starting points, these regions experience faster improvements in "people's well-being" indicators, thus achieving higher scores in the development outcome dimension. The fourth type is the "low input, low output" foundational cultivation type, represented by provinces including Liaoning, Guizhou, Heilongjiang, and Gansu. These regions lack sufficient inputs in factor supply, technological innovation, and industrial transformation, making it difficult to form the core driving force for NQPF development, which in turn constrains industrial efficiency improvement and the release of development outcomes. The results indicate that NQPF cannot "grow naturally" but are highly dependent on factor agglomeration capacity and institutional environment support.

6. Conclusion and Implications

This paper innovatively constructs an evaluation system for NQPF based on the "driving pathways–development outcomes" framework. Taking 31 provincial-level administrative regions in China as research samples, the study draws on multi-source statistical data to conduct an empirical measurement of the development level of new quality productive forces and to classify regional development types. The findings reveal: The findings reveal pronounced regional gradient differences, forming four echelon tiers (leading, active, steady, and emerging), with significant heterogeneity in development paths and dominant driving factors across regions. Moreover, the coupling coordination between driving pathways and development outcomes is generally low nationwide, with marked disparities in input-output conversion efficiency. The comprehensive measurement results indicate that the development of NQPF at the provincial level in China has already exhibited significant characteristics of head-agglomeration and regional differentiation. Further analysis of the internal structure of driving pathways reveals that the formation mechanisms of NQPF differ markedly across leading regions. At the regional level, a tailored, category-based policy approach should be adhered to, with precise efforts targeting the structural shortcomings of different regional types.

Looking ahead to the 15th Five-Year Plan period, developing new quality productive forces according to local conditions requires the following: First, given the cognitive bias inherent in traditional linear thinking—that "high input inevitably leads to high output"—it is necessary to reconstruct the assessment orientation of local governments and relevant departments. Second, the reliance on single input indicators such as the share of R&D expenditure and fixed-asset investment should be appropriately reduced, while greater weight should be assigned to indicators reflecting efficiency and transformation capacity. Third, it is essential to bridge the "valley of death" and advance the marketization mechanism for innovation achievements by establishing a full-chain, market-oriented acceleration mechanism. Fourth, fault-tolerance mechanisms should be strengthened to accommodate the diversity and uncertainty of new quality productive forces pathways. Given that the development of new quality productive forces is characterized by unclear technological trajectories and unverified business models, policy design must remain open and iterative. A fault-tolerance and trial-and-error

mechanism in the form of "policy sandboxes" or "innovation experimental zones" should be established. The government's role is not merely that of a factor allocator, but more importantly, that of a creator and enabler of the innovation ecosystem.

This study has the following limitations: First, constrained by data availability, certain dynamic indicators reflecting the new forms and models of new quality productive forces have not yet been incorporated into the statistical system. This study employs accessible publicly available proxy indicators for measurement, which may introduce measurement bias. Second, although the indicator system has been scientifically selected through the Delphi method and the Analytic Hierarchy Process, the selection of indicators itself cannot be entirely free from subjectivity, and certain potentially important variables may not have been captured. These aspects will be further refined in future research.

In summary, the essence of cultivating NQPF lies in promoting the formation of a virtuous cycle among factor input, achievement transformation, and development outcomes through institutional innovation and efficiency improvement. Whether optimizing top-level design, promoting category-based policies, or strengthening universal institutional supply, the core objective remains reducing transformation friction, improving allocation efficiency, and enabling innovation achievements to more fully translate into industrial competitiveness and public well-being.

References

[1] Bruna F. Market potential, spatial theories and spatial trends. Spatial Economic Analysis, 2024; 19(4), 527-548.

[2] Obschonka M., Tavassoli S., Rentfrow P. J., Potter, J., and Gosling, S. D. Innovation and inter-city knowledge spillovers: Social, geographical, and technological connectedness and psychological openness. Research Policy, 2023; 52(8), 104849.

[3] Israel, E. and Salinger, E. Forms of Innovation and the Core-Periphery Divide:(Non-) Technological Innovation and Geo‐Remoteness in Israel. Growth and Change, 2025; 56(1), e70019.

[4] Liu, H. and Li, X. How digital technology can improve new quality productive forces? —— Perspective of total factor agricultural carbon productivity. Journal of Asian Economics, 2025; 98, 101921.

[5] Leigh, N. G. and Kraft, B. R. Emerging robotic regions in the United States: insights for regional economic evolution. Regional Studies, 2018; 52(6), 804-815.

[6] Shi, X. T. and Yu, Z. Y. Measurement of new quality productive forces level, regional differences, and synergy degree of composite system. Science & Technology Progress and Policy, 2025; 42(7), 9–20.

[7] Liu, Z., Zhang, H., Guo, C. Y. and Yang, Y. New quality productive forces enabling high-quality development: Mechanism, measurement, and empirical analysis. Sustainability, 2025; 17(18), 8146.

[8] He, X. G. and Guo, X. B. Artificial intelligence technology innovation and firm total factor productivity. Economic and Management Research, 2026; 47(3), 144–160.

[9] Shang, S. and Feng, L. The effect of digitalization on urban green total factor productivity: Empirical evidence from China. Environment, Development and Sustainability, 2024; 6, 5013.

[10] Jiang, Y. M. and Qiao, Z. Y. Construction of an evaluation indicator system for the development of new quality productive forces. Reform of Economic System, 2024; (3), 5–15.

[11] Xu, A. T., Dai, Y. J. and Liu, Z. China's provincial new quality productive forces: Statistical measurement, dynamic evolution, and correlation network. Statistical Research, 2025; 42(10), 3–18.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261127

Configurational Pathways of Digital Transformation in New Energy Vehicle Manufacturing Enterprises

Jingrou Geng[1] and Yibing Tang
School of Management Science and Engineering, China Jiliang University, Hangzhou 310018, Zhejiang, China

Abstract. Against the backdrop of China's "Dual Carbon" strategy and the digital economy, the new energy vehicle (NEV) industry is undergoing a profound transformation featuring deep integration of electrification, intelligence, and connectivity. Digital transformation has become critical for NEV manufacturers seeking long-term competitiveness; however, due to high complexity and the absence of mature precedents, firm-level outcomes vary markedly. To explore the multiple driving mechanisms behind high-level digital transformation, this study integrates the Technology-Organization-Environment (TOE) framework with fuzzy-set qualitative comparative analysis (fsQCA). Using a panel of 31 A-share listed NEV firms (310 firm-year observations, 2014-2024), we examine six antecedent conditions: digital intangible assets (DA), R&D intensity (RD), executive compensation (MS), technical talent density (TC), government subsidies (GS), and market competition (IC). We find that: (1) high-level digital transformation results from the concurrent synergy of multiple conditions, and no single condition is individually necessary; (2) three equifinal pathways are identified-technology-organization dual internal-driven, organization-environment market-driven, and organization-environment policy-leveraged-jointly covering 49.5% of high-level transformation cases; and (3) executive compensation acts as a "meta-condition" across all pathways, technical talent density and market competition form a stable linkage, and government subsidies and R&D investment exhibit a cross-configurational substitution effect. These findings suggest that NEV firms should prioritize executive incentives and talent reserves, while policymakers should shift from universal subsidies to targeted, capability-contingent support.

Keywords. Digital transformation; New energy vehicle; TOE framework; fsQCA; Configurational pathway

1. Introduction

Driven by China's "Dual Carbon" strategy and the rapid expansion of the digital economy, the new energy vehicle (NEV) industry is undergoing profound structural transformation. As electrification, intelligence, and connectivity become increasingly integrated, digital transformation has emerged as a critical source of competitive advantage for NEV manufacturers[1]. However, owing to the complexity of transformation processes and the absence of established implementation models, firms exhibit substantial variation in their transformation outcomes[2].

[1] Corresponding Author, Jingrou Geng; Email: 1967211079@qq.com

Traditional linear regression approaches assume independent and additive causal effects, which cannot capture the complex, conjunctural, and equifinal nature of digital transformation. Digital transformation simultaneously involves technological endowments, organizational governance, and environmental pressures. fsQCA overcomes this limitation by examining multiple conjunctural causation and equifinality, making it uniquely suited to this context.

1.1. Research gap and contributions

Existing research on digital transformation has predominantly focused on general manufacturing sectors or examined the independent effects of individual antecedents[3]. Although these studies provide valuable insights, they offer limited understanding of the unique transformation dynamics of the NEV industry. Compared with traditional manufacturing sectors, NEV firms operate in a policy-intensive environment[4], face rapid technological transitions associated with electrification, and experience the simultaneous evolution of product and digital innovation. To address this gap, this study integrates the TOE framework with fsQCA to examine configurational pathways of digital transformation in NEV firms. We make three contributions: (1) revealing executive compensation as a universal "meta-condition"; (2) identifying a cross-configurational substitution effect between government subsidies and R&D investment; and (3) applying TOE-fsQCA to a panel NEV dataset, an approach rarely used in this context.

2. Theoretical Framework

2.1. TOE Framework

Proposed by Tornatzky & Fleischer[5], the TOE framework posits that technological innovation adoption results from interactions among technological endowments, organizational resources and governance, and environmental pressures (policy, market, ecosystem). Its three-dimensional structure reflects the technology-intensive, organizationally collaborative, and policy-driven characteristics of NEV digital transformation, providing the analytical basis for condition selection in fsQCA[3].

2.2. Theoretical Extensions

We draw on three complementary perspectives. First, dynamic capability theory holds that firms must develop sensing, seizing, and reconfiguring capabilities for digital innovation[6]; in the NEV context, RD and DA jointly reflect such capabilities. Second, the resource orchestration perspective emphasizes that effective bundling and leveraging—not mere possession—of resources drives performance[7]. TC and MS are thus interpreted as organizational enablers of resource mobilization. Third, strategic leadership theory highlights that top management incentive alignment is critical for initiating large-scale transformation[8]; MS operates as a governance mechanism aligning managerial incentives with long-term digital investment.

2.3. Configurational Propositions

P1 (Technology-Organization Internal-Driven): Firms with high DA, high RD, and strong MS are likely to achieve high-level digital transformation through internally driven pathways, reducing reliance on external policy support. P2 (Organization-Environment Market-Driven): Firms with strong TC, proactive top management (MS), and intense market competition (IC) can achieve transformation through market pressure-driven pathways. P3 (Organization-Environment Policy-Leveraged): For firms with weaker internal RD, adequate MS combined with GS and IC can substitute for R&D investment and enable high-level transformation.

2.4. Condition Operationalization

Technological dimension: DA (ratio of digital intangible assets to total intangible assets [9]) and RD (R&D expenditure / operating revenue[10]). Organizational dimension: MS (ln of average total compensation of top-three executives[11]) and TC (technical personnel / total employees[12]). Environmental dimension: GS (ln of annual government subsidies received[4]) and IC measured by the industry price-cost margin (Lerner index)—a lower value indicates more intense competition[13][14]. We prefer the Lerner index over HHI as the latter cannot accurately reflect competitive conditions under specific market structures[15][16]. The outcome variable DTL follows Zhao et al.[17], using keyword frequency in annual reports across four dimensions (digital technology application, internet business models, intelligent manufacturing, modern information systems), synthesized via the entropy method. DTL denotes the continuous pre-calibration measure; after applying the direct calibration method, firm-year observations with a fuzzy membership score exceeding 0.5 constitute the outcome set referred to throughout as high-level digital transformation—the two terms are therefore analytically distinct and are used consistently in their respective roles in all subsequent sections. We acknowledge that text-based DTL measurement may partially reflect disclosure intensity rather than actual operational transformation. However, China's mandatory CSRC disclosure regulations and independent audit requirements create structural disincentives for material overstatement; the multi-dimensional entropy-weighted synthesis across four distinct keyword categories further reduces single-axis inflation risk. This limitation is explicitly noted in the Conclusion.

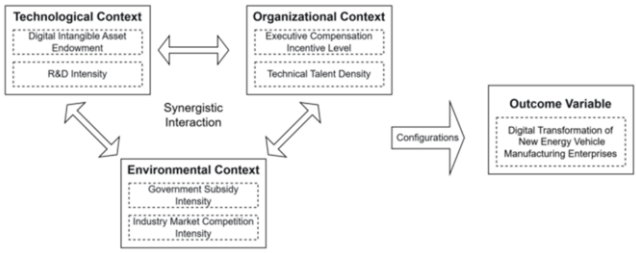


Figure 1. Configurational model of digital transformation for NEV manufacturers.

3. Research Design

3.1. Data Collection

We collected A-share listed NEV manufacturing companies from 2014–2024 from corporate annual reports, the CSMAR database, and the State Intellectual Property Office. After excluding ST/*ST firms and those with severe missing data (imputed via linear interpolation), the final sample comprised 31 firms and 310 firm-year observations. All antecedent conditions were lagged by one period to mitigate reverse causality (conditions: 2014–2023; outcome: 2015–2024).The one-period lag structure also mitigates impression management bias: condition variables in year t-1 are derived from audited financial statements, headcount records, patent registries, and government transfer data—sources independent of annual report narratives—ensuring that a firm's disclosure strategy in year t cannot retroactively affect its explanatory conditions. Cross-source triangulation across the CSMAR database, State Intellectual Property Office, and annual reports further reduce the risk that any single disclosure channel systematically determines both conditions and outcome.

3.2. Calibration

We applied the direct calibration method using the 95th, 50th, and 5th percentiles as anchors for full membership (0.95), crossover point (0.50), and full non-membership (0.05), respectively[18]. Membership scores of exactly 0.5 were recorded to 0.501 to retain cases at the crossover point. Table 1 reports the anchors.

Table 1. Calibration Anchors for Each Variable

Variable	Full Membership	Crossover Point	Full Non-Membership
DA	0.1603	0.048	0.004
RD	7.137	4.275	1.704
MS	15.661	14.557	13.517
TC	0.359	0.168	0.054
GS	19.971	15.897	11.050
IC	0.080	0.058	0.045
DTL	0.020	0.007	0.002

4. Data Analysis

4.1. Necessity Analysis

We first tested whether any single condition is necessary for high-level digital transformation (threshold: consistency > 0.9[19]). As shown in Table 2, no condition or its negation exceeds this threshold, confirming that high-level digital transformation requires the concurrent synergy of multiple conditions rather than any single factor, consistent with Propositions 1–3.

Table 2. Necessity Analysis Results of Conditional Variables

Conditional Variable	High Digital Transformation Level		Low Digital Transformation Level	
	Consistency	Coverage	Consistency	Coverage
DA	0.598	0.620	0.580	0.680
~DA	0.6918	0.593	0.675	0.655
RD	0.7378	0.679	0.613	0.639
~RD	0.608	0.582	0.692	0.748
MS	0.726	0.692	0.569	0.613
~MS	0.594	0.550	0.714	0.746
TC	0.682	0.666	0.585	0.646
~TC	0.637	0.576	0.697	0.712
GS	0.655	0.612	0.648	0.684
~GS	0.662	0.625	0.632	0.674
IC	0.689	0.749	0.501	0.615
~IC	0.646	0.534	0.795	0.743

4.2. Configurational Analysis

Following Kumar S et al.[20], we set the raw consistency threshold at 0.8 and PRI consistency at 0.7. Given the 310-observation panel, the case frequency threshold was set to 3, consistent with Fiss[21]. Table 3 reports three configurations, all with individual consistencies above 0.90. The overall solution consistency is 0.892 and coverage is 0.496. Core conditions (●) appear in the parsimonious solution; peripheral conditions (•) appear only in the intermediate solution; ⊗ denotes core absence; blank cells indicate no systematic relationship.

Table 3. Configurations for high digital transformation in NEV enterprises

Antecedent condition	High-Level Digital Transformation		
	Technology-Organization Dual Internal-Driven Path	Organization-Environment Market-Driven Path	Organization-Environment Policy-Leveraged Path
DA	●		
RD	●	•	
MS	●	●	●
TC		●	●
GS	⊗		•
IC		●	●
Consistency	0.910	0.904	0.905
Raw coverage	0.319	0.401	0.348
Unique coverage	0.0640	0.028	0.031
Overall consistency		0.892	
Overall coverage		0.495	

Note: ● = core condition present; • = peripheral condition present; ⊗ = core condition absent; blank = no influence.

Configuration 1: Technology-Organization Dual Internal-Driven (DA·RD·MS·~GS). Consistency = 0.910; raw coverage = 0.319. This pathway relies entirely on internal resources and governance: high DA, high RD, and strong MS jointly drive transformation without external policy support. The core absence of GS signals self-sustaining transformation capability rather than policy dependency, consistent with dynamic capability theory —firms accumulating superior technological assets and aligned governance generate sufficient internal transformation momentum. Weichai Power exemplifies this pathway: 2023 R&D intensity ≈ 6.5% (industry-leading) and extensive software copyrights in smart factories and digital twins. This extends prior fsQCA findings in broader manufacturing contexts [3]to the NEV institutional environment. This finding provides support for Proposition P1.

Configuration 2: Organization-Environment Market-Driven (RD·MS·TC·IC). Consistency = 0.904; raw coverage = 0.401 (most explanatory pathway). Intense market competition (IC) creates external pressure; ample technical talent (TC) and executive commitment (MS) constitute organizational responsiveness. RD as a peripheral condition provides auxiliary absorptive capacity without being a core prerequisite, consistent with the resource orchestration perspective[7]: the synergy of talent, managerial commitment, and competitive pressure—not any single resource—enables transformation. SAIC Motor illustrates this: despite fierce competition, large R&D teams and high management incentives enabled its subsidiary SAIC Maxus to become a national smart manufacturing benchmark. Compared to studies emphasizing R&D as the primary driver, this pathway highlights the under-appreciated role of market pressure in triggering organizational mobilization. This finding provides support for Proposition P2.

Configuration 3: Organization-Environment Policy-Leveraged (MS·TC·GS·IC). Consistency = 0.905; raw coverage = 0.348. This pathway shares the same core conditions as Configuration 2 (MS, TC, IC) but substitutes GS for RD as the peripheral condition—revealing a policy-complementarity pathway where subsidies compensate for weaker internal R&D. Chang'an Automobile illustrates this: 17,000-person technical R&D team, 600+ core intelligent and low-carbon technologies, and ≥5% R&D spending ratio (solid organizational base), combined with substantial NEV and smart manufacturing subsidies as a central state-owned enterprise subsidiary. Subsidies are thus most effective for firms with strong organizational capabilities but limited R&D endowments, with direct implications for targeted subsidy design. This finding provides support for Proposition P3.

Three cross-pathway regularities emerge. First, MS appears as a core condition in all three configurations—acting as a "meta-condition" indispensable for digital transformation regardless of other conditions. This is consistent with the strategic leadership and upper echelons literature[8]: executive incentive alignment reduces principal-agent costs and constitutes a prerequisite for long-term, high-uncertainty digital investment. The universality of MS across diverse technological and environmental contexts extends prior findings. Second, TC and IC form a stable core combination in Configurations 2 and 3, suggesting that competitive pressure and human capital co-evolve as organizational-environmental complements in driving transformation, echoing the dynamic capability perspective's emphasis on external sensing and internal reconfiguring [6]. Third, a cross-configurational substitution effect exists between GS and RD: Configuration 2 relies on RD as peripheral, while Configuration 3 relies on GS—given identical core conditions, both achieve equivalent high-level transformation. This substitution effect is consistent with the innovation economics literature and has

direct implications for differentiated subsidy policy: public support is most effective when targeting firms with strong organizational capabilities but weak R&D endowments.

4.3. Robustness Check

To test the robustness of our configurational analysis results, this study followed established practices by raising the consistency threshold. The original consistency threshold was raised from 0.8 to 0.85, while keeping PRI consistency and other parameters unchanged, and the configurational analysis was re-conducted. The results show that, under the new threshold, the number of identified configurations, the composition of core conditions, and the overall coverage remain highly consistent with the original results, indicating strong robustness of our findings.

5. Conclusion

This study applied TOE-fsQCA to a panel of 31 NEV firms (310 firm-year observations) to examine the configurational drivers of high-level digital transformation. Four key findings emerge: (1) high–level digital transformation results from the concurrent synergy of multiple conditions; no single necessary condition exists; (2) three equifinal pathways are identified–technology–organization dual internal-driven, organization-environment market-driven, and organization-environment policy-leveraged—with overall coverage of 49.5%; (3) executive compensation incentive is the universal "meta-condition" across all pathways, consistent with strategic leadership theory ; and (4) GS and RD exhibit a cross-configurational substitution effect, with direct implications for differentiated subsidy design.

5.1. Theoretical Implications

this study extends the TOE framework by integrating dynamic capability[6], resource orchestration[7], and strategic leadership perspectives, contributing to configurational digital transformation research by demonstrating equifinality and asymmetric causation in the NEV sector. Compared to prior linear studies[1][2]. The identification of executive compensation as a universal meta-condition extends the upper echelons literature[8] to digital transformation contexts, while the GS-RD substitution effect enriches the innovation economics literature on public-private R&D complementarity.

5.2. Managerial Implications

Our findings reveal that transformation drivers cannot be understood in isolation. For practice, NEV firms should prioritize executive incentives and talent reserves, then select pathways fitting their technological and environmental context. Policymakers should build differentiated support systems—providing targeted subsidies to firms with strong organizational capabilities but weak R&D, while shifting to outcome-based incentives for internally driven firms.

5.3. Limitations and Future Research

First, the sample size (N=31) is a limitation; future research should expand coverage, adopt longitudinal QCA designs, and extend to cross-national NEV contexts. Second, the text-based DTL measure may partially capture disclosure practices rather than actual transformation depth; future research should triangulate with operational measures such as IT investment ratios or patent citation networks when such data become available.

References

[1] Lou Z, Gao N, Lu M. The impact of enterprise digital transformation on low-carbon supply chains: empirical evidence from China. Sustainability, 2024, 16(18): 8242. doi: 10.3390/su16188242.

[2] Liu J, Wang Z, Li C, Xu R. Exploring pathways to digital transformation: fsQCA analysis based on the AMO framework. PLOS ONE, 2024, 19(12): e0315249. doi: 10.1371/journal.pone.0315249.

[3] Miao Z, Zhao G. Configurational paths to the green transformation of Chinese manufacturing enterprises: a TOE framework based on the fsQCA and NCA approaches. Scientific Reports, 2023, 13: 19181. doi: 10.1038/s41598-023-46454-9.

[4] Wang S, Li X, Li Z, Ye Y. The effects of government support on enterprises' digital transformation: evidence from China. Managerial and Decision Economics, 2023, 44(5): 2520–2539. doi: 10.1002/mde.3831.

[5] Tornatzky L G, Fleischer M, Chakrabarti A K. The processes of technological innovation. Massachusetts: Lexington Books, 1990.

[6] Teece D J. Explicating dynamic capabilities: the nature and microfoundations of (sustainable) enterprise performance. Strategic Management Journal, 2007, 28(13): 1319-1350. doi: 10.1002/smj.640.

[7] Sirmon D G, Hitt M A, Ireland R D. Managing firm resources in dynamic environments to create value: looking inside the black box. Academy of Management Review, 2007, 32(1): 273-292. doi: 10.5465/amr.2007.23466005.

[8] Hambrick D C, Mason P A. Upper echelons: the organization as a reflection of its top managers. Academy of Management Review, 1984, 9(2): 193-206. doi: 10.5465/ambpp.1982.4976402.

[9] Li N, Wang X, Wang Z, Luan X. The impact of digital transformation on corporate total factor productivity. Frontiers in Psychology, 2022, 13: 1071986. doi: 10.3389/fpsyg.2022.1071986.

[10] Cohen W M, Levinthal D A. Absorptive capacity: a new perspective on learning and innovation. Administrative Science Quarterly, 1990, 35(1): 128-152. doi: 10.2307/2393553.

[11] Zhou Y, Wang H, Lan H. Why and how executive equity incentive influences digital transformation. Technology Analysis & Strategic Management, 2024, 36(12): 4217-4231. doi: 10.1080/09537325.2023.2250012.

[12] Peng D Y, Zhang B K. Research on the impact of enterprise digital transformation on labor employment in China. Social Governance Review, 2025, (5): 71-93. doi: 10.3969/j.issn.2096-0786.2025.05.007.

[13] Zhou X F, Zhou Q L. Product market power, industry competition, and corporate earnings management: based on the empirical evidence of Chinese listed companies. Accounting Research, 2014, (8): 60-66.

[14] Cremers K J M, Nair V B, Peyer U. Takeover defenses and competition: the role of stakeholders. Journal of Empirical Legal Studies, 2008, 5(4): 791-818. doi: 10.1111/j.1740-1461.2008.00141.x.

[15] Song Q, Liu Y H. Market competition, R&D investment and innovation output of technology-based SMEs. China Soft Science, 2021, (10): 182-192.

[16] Chen B L, Zhong H Y. The 'double-edged sword' effects of the geographic concentration of institutional investors on corporate innovation. Journal of Industrial Engineering and Engineering Management, 2025, 39(6): 134-148.

[17] Zhao C Y, Wang W C, Li X S. How does digital transformation affect the total factor productivity of enterprises? Finance & Trade Economics, 2021, 42(7): 114-129.

[18] Guedes M J, da Conceicao Goncalves V, Soares N, et al. UK evidence for the determinants of R&D intensity from a panel fsQCA. Journal of Business Research, 2016, 69(11): 5431-5436. doi: 10.1016/j.jbusres.2016.04.150.

[19] Ragin C C, Fiss P C. Net effects analysis versus configurational analysis: an empirical demonstration. Redesigning Social Inquiry: Fuzzy Sets and Beyond, 2008: 190–212.

[20] Kumar S, Sahoo S, Lim W M, et al. Fuzzy-set qualitative comparative analysis (fsQCA) in business and management research: a contemporary overview. Technological Forecasting and Social Change, 2022, 178: 121599. doi: 10.1016/j.techfore.2022.121599.

[21] Fiss P C. Building better causal theories: a fuzzy set approach to typologies in organization research. Academy of Management Journal, 2011, 54(2): 393–420. doi: 10.5465/AMJ.2011.60263120.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)

© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261128

Emission Reduction and Pricing Strategies in Low-Carbon Supply Chains under Information Asymmetry and Risk Aversion

Yuxin Li [a,1] and Yuxiang Yang[a]
[a] *China Jiliang University, China*
ORCiD ID: Yuxin Li https://orcid.org/0009-0003-6635-8822

Abstract. Under carbon taxation, firms' abatement behavior is shaped by cost pressures, the information environment, and risk preferences. Yet most existing studies focus on single supply chains, with limited attention to how information asymmetry and risk aversion jointly affect manufacturers' abatement and pricing under horizontal competition. To fill this gap, this paper models two competing supply chains under carbon taxation and technology spillovers, using a mean-variance Stackelberg game to derive equilibria under centralized and decentralized decision-making, with and without information sharing. Results show that risk aversion induces a defensive "high pricing, low abatement" strategy—most pronounced without information sharing and nearly eliminated under information sharing, suggesting that abatement transparency substitutes for risk tolerance. Beyond a threshold, the tax suppresses investment and exhibits an inverted-U relationship; spillover effects depend critically on the information environment. These findings assist firms in choosing between integration and sharing, and inform policymakers in tailoring carbon taxes to industry information transparency.

Keywords. Dual competing supply chain; emission reduction; carbon tax; technology spillover; information sharing; risk aversion

1. Introduction

Carbon taxes, as a primary policy instrument under the global "dual carbon" targets, have reshaped firms' abatement costs and pricing logic. In carbon-constrained markets, inter-chain competition and technology spillovers are prevalent. While spillovers reduce industry-wide abatement costs, they also induce free-riding that undermines firms' independent incentives for emission reduction. However, most existing studies on carbon taxes and spillovers assume risk-neutral decision-makers—a premise inconsistent with observed managerial behavior. When confronting uncertainty, risk-averse firms adopt conservative pricing and cautious abatement strategies to hedge against profit volatility. Overlooking this behavioral trait may lead to systematic overestimation of abatement investment and underestimation of equilibrium prices.

The role of risk aversion in low-carbon supply chain decisions has received increasing scholarly attention. Nevertheless, existing work remains confined to single-chain settings, leaving unaddressed how risk-averse decisions are jointly shaped by rivals' actions and spillover externalities under two-chain horizontal competition with cross-chain spillovers. Information asymmetry introduces additional complexity: competing manufacturers often conceal their abatement efforts, forcing rivals to form

[1] Corresponding Author: Yuxin Li, Email: lliyxin@163.com

beliefs rather than rely on observed data. The literature, however, focuses predominantly on demand-side or vertical information, with limited scrutiny of cross-chain low-carbon investment asymmetry.

These three constraints—risk preference, information structure, and inter-chain competition—are not mutually independent. Existing studies treat them separately, thereby overlooking their interactions. To address this gap, this paper examines emission reduction and pricing under risk aversion and information asymmetry within a framework of carbon taxation and technology spillovers, revealing how risk aversion and information structure jointly determine low-carbon investment and supply chain performance.

2. Literature Review

This section reviews three strands of literature: abatement decisions under carbon taxes and spillovers, risk aversion in low-carbon supply chains, and information asymmetry and sharing.

On carbon taxes and spillovers, Liu et al.[1] demonstrate that carbon taxes promote abatement by incentivizing green investment while curbing free-riding among competitors. Wang et al.[2] find that horizontal competition and negative spillovers inhibit abatement. Jiang et al.[3] report that spillover direction differentially affects equilibrium outcomes. Wu et al.[4] find that high-intensity spillovers suppress green R&D, whereas low-intensity spillovers encourage it, and that manufacturer altruism enhances investment. Yet all these studies assume risk neutrality.

With respect to risk aversion, the mean-variance framework has been widely adopted[5]. Feng et al.[6] demonstrate that retailer risk aversion alters pricing and abatement decisions under carbon taxes. Chen et al.[7] confirm that risk-averse manufacturers cut abatement and pass costs through higher wholesale prices. Chen et al.[8] report that stronger risk aversion lowers wholesale and retail prices and product greenness, and that cost-sharing contracts enhance profits under weak competition. Barman and Sana[9] observe that risk aversion reduces uncertainty but diminishes overall profit. These studies, however, are confined to single supply chains.

With respect to information asymmetry, Ma et al. [10] find that concealing green preference information improves environmental benefits but harms the less-informed party. Hu et al. [11] show that demand information structure alters profit distribution. Xia and Niu [12] find that information asymmetry distorts decisions and that screening contracts mitigate this distortion. Yet these studies give limited attention to horizontal low-carbon investment information between competing supply chains.

Three limitations emerge: the carbon tax and spillover literature assumes risk neutrality; the risk aversion literature is confined to single chains; and the information asymmetry literature neglects horizontal investment information. These limitations are not independent—risk preference determines the value of information, information structure delimits observability, and inter-chain competition transmits the effects of both. Whether existing findings hold when all three features coexist is an open question. To fill this gap, this paper develops a unified framework to address four research questions: (1) How does risk aversion affect low-carbon investment and pricing under information non-sharing? (2) Does the effectiveness of information sharing depend on the degree of risk aversion? (3) How do carbon tax, spillovers, and competition interact with information structure and risk preference? (4) What are the optimal firm strategies and policy implications under different configurations of these factors?

3. Model construction and assumptions

We consider two competing two-echelon supply chains, each consisting of a manufacturer and a retailer producing substitutable low-carbon products. The government levies a carbon tax per unit of emissions. Technology spillovers allow one manufacturer's investment to affect the competitor's emissions. Information asymmetry takes two forms: under sharing, both parties' investments are common knowledge; under non-sharing, manufacturers cannot observe the competitor's investment and treat it as a random variable. Risk-averse manufacturers maximize mean-variance utility. The Stackelberg game proceeds as follows: manufacturers first decide whether to share information, then choose investment levels and wholesale prices; retailers then set retail prices. Four scenarios are analyzed—centralized-sharing (CS), centralized-non-sharing (CN), decentralized-sharing (DS), and decentralized-non-sharing (DN).

Assumption 1: Following Chen et al.[13], the inverse demand function is specified as $p_i = A - q_i - \beta q_j, i = 1, j = 3 - i$, where A is base market demand and $\beta \in [0,1]$ is the competition intensity. Low-carbon investment incurs a quadratic cost $\frac{1}{2}ht_i^2$, where $h(h > 0)$ is the cost coefficient. Unit production cost is normalized to zero.

Assumption 2: Under information non-sharing, a manufacturer cannot directly observe the competitor's low-carbon investment and must instead form a belief based on historical experience, treating it as a random variable $t_j \sim N(\bar{t}, \sigma_j^2)$ .Under information sharing, both parties observe each other's investment level t_j without bias.

Assumption 3: Manufacturers exhibit risk-averse behavior. Their risk preference is characterized by a mean-variance utility function $U = E(\pi_m) - \eta\sqrt{Var(\pi_m)}$ [5], where $E(\pi_m)$ denotes the expected profit, $Var(\pi_m) = E\left[\pi_m - E(\pi_m)\right]^2$ denotes the profit variance, and η is the risk aversion coefficient, with larger values indicating higher risk aversion.

Assumption 4: The government levies a fixed carbon tax $\tau(\tau > 0)$ per unit of emissions. With initial emissions normalized to 1 and spillover coefficient $s \in (0,1)$, the final emission per unit is $e_i = 1 - t_i - st_j$, yielding a carbon tax cost of $\tau(1 - t_i - st_j)$ [3].

Assumption 5: The two supply chains are perfectly symmetric, and the symmetric Nash equilibrium is adopted. To ensure the existence and economic relevance of equilibrium solutions, the parameters satisfy $0 < \eta < \frac{A - \tau}{s\sigma\tau}, A > \tau, h \geq \frac{(1+s)^2\tau^2}{1+\beta}$.

4. Model solving

This section formulates and solves the dual-chain competition models under centralized and decentralized decision-making with different information structures.

4.1 Decentralized-Non-sharing (DN)

Under DN, the manufacturer's and retailer's profit functions are: $\pi_{mi}^{DN} = \left(\omega_i - \tau\left(1 - t_j - st_j\right)\right)q_i - \frac{1}{2}ht_i^2$, $\pi_{ri}^{DN} = (p_i - \omega_i)q_i$. Since information is not shared, the competitor's low-carbon investment is a random variable. The manufacturer's mean-variance utility is $U_i^{DN} = \left(\omega_i - \tau\left(1 - t_j - s\bar{t_j}\right)\right)q_i - \frac{1}{2}ht_i^2 - \eta\sqrt{\tau^2 s^2 q_i^2 \sigma^2}$.

Solving via backward induction yields the equilibrium outcomes. The low-carbon

investment level is $t_i^{DN*}=\frac{(1+s)\tau(-A+\tau+s\eta\sigma\tau)}{-4h(1+\beta)+(1+s)^2\tau^2}$, and the wholesale price is $\omega_i^{DN*}=\frac{2Ah(1+\beta)+2h(1+\beta)(1+s\eta\sigma)\tau-A(1+s)^2\tau^2}{4h(1+\beta)-(1+s)^2\tau^2}$. The retail price is $p_i^{DN*}=A-\frac{h(1+\beta)(A-(1+s\eta\sigma)\tau)}{4h(1+\beta)-(1+s)^2\tau^2}$. Correspondingly, the retailer's profit is $\pi_{ri}^{DN*}=\frac{h^2(1+\beta)(-A+\tau+s\eta\sigma\tau)^2}{\left(-4h(1+\beta)+(1+s)^2\tau^2\right)^2}$, the manufacturer's profit is $\pi_{mi}^{DN*}=\frac{h^2(-A+\tau+s\eta\sigma\tau)^2}{8h(1+\beta)-2(1+s)^2\tau^2}$, and the total supply chain profit is $\pi_i^{DN*}=\frac{h(-A+\tau+s\eta\sigma\tau)^2\left(6h(1+\beta)-(1+s)^2\tau^2\right)}{2\left(-4h(1+\beta)+(1+s)^2\tau^2\right)^2}$,for $i=1,2$.

4.2 Decentralized-Sharing (DS)

Under DS, manufacturers observe each other's investments and maximize their own profits. The profit functions are: $\pi_{ri}^{DS}=(p_i-\omega_i)q_i$, $\pi_{mi}^{DS}=\left(\omega_i-\tau\left(1-t_j-st_j\right)\right)q_i-\frac{1}{2}ht_i^2$.

Solving via backward induction yields the equilibrium outcomes. The low-carbon investment level is $t_i^{DS*}=-\frac{(1+s)\tau(A-\tau)}{-4h(1+\beta)+(1+s)^2\tau^2}$, and the wholesale price is $\omega_i^{DS*}=\frac{2Ah(1+\beta)+2h(1+\beta)\tau-A(1+s)^2\tau^2}{4h(1+\beta)-(1+s)^2\tau^2}$. The retail price is $p_i^{DS*}=A-\frac{h(1+\beta)(A-\tau)}{4h(1+\beta)-(1+s)^2\tau^2}$. Correspondingly, the retailer's profit is $\pi_{ri}^{DS*}=\frac{h^2(1+\beta)(A-\tau)^2}{\left(-4h(1+\beta)+(1+s)^2\tau^2\right)^2}$, the manufacturer's profit is $\pi_{mi}^{DS*}=\frac{h^2(A-\tau)^2}{8h(1+\beta)-2(1+s)^2\tau^2}$, and the total supply chain profit is $\pi_i^{DS*}=\frac{h(A-\tau)^2\left(6h(1+\beta)-(1+s)^2\tau^2\right)}{2\left(-4h(1+\beta)+(1+s)^2\tau^2\right)^2}$,for $i=1,2$.

4.3 Centralized-Non-sharing (CN)

Under CN, the manufacturer and retailer act as a single integrated entity. The total supply chain profit is: $\pi_i^{CN}=\left(p_i-\tau\left(1-t_i-st_j\right)\right)q_i-\frac{1}{2}ht_i^2$. Since information is not shared, the competitor's low-carbon investment is a random variable. The mean-variance utility is $U_i^{CN}=\left(p_i-\tau\left(1-t_i-st_j\right)\right)q_i-\frac{1}{2}ht_i^2-\eta\sqrt{\tau^2s^2q_i^2\sigma^2}$.

Solving yields the equilibrium outcomes. The low-carbon investment level is $t_i^{CN*}=\frac{(1+s)(A-\tau)\tau}{-(1+s)^2\tau^2+2h\left(1+\beta+s^2\eta\sigma^2\tau^2\right)}$. The retail price is $p_i^{CN*}=A-\frac{h(1+\beta)(A-\tau)}{-(1+s)^2\tau^2+2h\left(1+\beta+s^2\eta\sigma^2\tau^2\right)}$. The total supply chain profit is $\pi_i^{CN*}=\frac{h(A-\tau)^2}{-2(1+s)^2\tau^2+4h\left(1+\beta+s^2\eta\sigma^2\tau^2\right)}$,for $i=1,2$.

4.4 Centralized-Sharing (CS)

Under CS, the competitor's low-carbon investment is directly observable. The total supply chain profit function is $\pi_i^{CS}=\left(p_i-\tau\left(1-t_i-st_j\right)\right)q_i-\frac{1}{2}ht_i^2$. Solving via first-order conditions yields the equilibrium outcomes. The low-carbon investment level

is $t_i^{CS*} = -\frac{(1+s)(A-\tau)\tau}{-2h(1+\beta)+(1+s)^2\tau^2}$, and the retail price is $p_i^{CS*} = A - \frac{h(1+\beta)(A-\tau)}{2h(1+\beta)-(1+s)^2\tau^2}$. The total supply chain profit is $\pi_i^{CS*} = \frac{h(A-\tau)^2}{4h(1+\beta)-2(1+s)^2\tau^2}$, for $i = 1,2$.

5. Model analysis

Proposition 1: Information sharing increases low-carbon investment under both structures: $t_i^{DS*} > t_i^{DN*}$ and $t_i^{CS*} > t_i^{CN*}$. Depending on the risk aversion coefficient, three rankings emerge: When $0 < \eta < \frac{1+\beta}{s^2\sigma^2\tau^2}$, $t_i^{CS*} > t_i^{CN*} > t_i^{DS*} > t_i^{DN*}$; When $\frac{1+\beta}{s^2\sigma^2\tau^2} < \eta < \eta_1$ or $\eta_2 < \eta < \frac{A-\tau}{s\sigma\tau}$, $t_i^{CS*} > t_i^{DS*} > t_i^{CN*} > t_i^{DN*}$; when $\eta_1 < \eta < \eta_2$, $t_i^{CS*} > t_i^{DS*} > t_i^{DN*} > t_i^{CN*}$, where $\eta_1 < \eta_2$ and $\eta_1, \eta_2 \in \left(\frac{1+\beta}{s^2\sigma^2\tau^2}, \frac{A-\tau}{s\sigma\tau}\right)$. For detailed derivations, please refer to the appendix[2].

When risk aversion is low, CN yields the highest investment. As risk aversion increases, DS overtakes CN, and DN may do so under certain conditions. Manufacturers should therefore align their coordination choices with risk preferences, prioritizing cross-chain information sharing when risk aversion is high. Policymakers may promote such sharing as a substitute for antitrust intervention in vertical integration.

Proposition 2: Comparing wholesale prices yields: $\omega_i^{DS*} < \omega_i^{DN*}$ .Information sharing lowers wholesale prices by eliminating the risk premium associated with carbon price and demand volatility. Retailers should therefore partner with manufacturers that engage in extensive information sharing to gain pricing flexibility. Manufacturers, in turn, can secure higher order volumes through lower wholesale prices—a strategy particularly effective when demand price elasticity is high.

Proposition 3: Comparing retail prices yields $p_i^{DS*} < p_i^{DN*}$ and $p_i^{CS*} < p_i^{CN*}$. Depending on the risk aversion coefficient, three rankings emerge: when $0 < \eta < \frac{1+\beta}{s^2\sigma^2\tau^2}$, $p_i^{CS*} < p_i^{CN*} < p_i^{DS*} < p_i^{DN*}$; when $\frac{1+\beta}{s^2\sigma^2\tau^2} < \eta < \eta_1$ or $\eta_2 < \eta < \frac{A-\tau}{s\sigma\tau}$, $p_i^{CS*} < p_i^{DS*} < p_i^{CN*} < p_i^{DN*}$; when $\eta_1 < \eta < \eta_2$, $p_i^{CS*} < p_i^{DS*} < p_i^{DN*} < p_i^{CN*}$.

Under information sharing, retail prices are lower than under non-sharing, and centralization reduces them further. When risk aversion is high, however, DS undercuts CN, indicating that information sharing achieves deeper price cuts than vertical integration under extreme risk. Retailers should therefore promote information sharing among manufacturers to reduce final prices, rather than relying on vertical integration.

Proposition 4: A comparison of profits clearly shows that $\pi_i^{DS*} > \pi_i^{DN*}$ and $\pi_i^{CS*} > \pi_i^{CN*}$, then depending on the risk aversion coefficient, the profit rankings are as follows: when $0 < \eta < \frac{1+\beta}{3s^2\sigma^2\tau^2}$, $\pi_i^{CN*} > \pi_i^{DS*}$, giving $\pi_i^{CS*} > \pi_i^{CN*} > \pi_i^{DS*} > \pi_i^{DN*}$; when $\frac{1+\beta}{3s^2\sigma^2\tau^2} < \eta < \frac{A-\tau}{s\sigma\tau}$, $\pi_i^{CN*} < \pi_i^{DS*}$, and the ranking depends on the cost coefficient h :if $0 < h < h_1$ or $h > h_2$, then $\pi_i^{CS*} > \pi_i^{DS*} > \pi_i^{CN*} > \pi_i^{DN*}$; if $h_1 < h < h_2$, then $\pi_i^{CS*} > \pi_i^{DS*} > \pi_i^{DN*} > \pi_i^{CN*}$.

Both mechanisms improve profitability. CN performs best when abatement costs are low; DS/DN dominates when costs are high. Manufacturers should accordingly choose vertical integration for low costs and cross-chain sharing for high costs. Retailers in high-cost industries should participate actively in cross-chain sharing to

[2] For the derivation and verification of the relevant numerical values in the main text, please refer to the appendix: https://www.kdocs.cn/l/ctzxlL5YXUAK

capture the gains.

Proposition 5: presents the partial derivatives of the equilibrium solutions with respect to carbon tax τ : $\frac{\partial p_i^{DN*}}{\partial \tau}>0,\frac{\partial p_i^{DS*}}{\partial \tau}>0,\frac{\partial p_i^{CN*}}{\partial \tau}>0,\frac{\partial p_i^{CS*}}{\partial \tau}>0;\frac{\partial t_i^{DN*}}{\partial \tau}>0,\frac{\partial t_i^{DS*}}{\partial \tau}>0,\frac{\partial t_i^{CS*}}{\partial \tau}>0;\frac{\partial \omega_i^{DN*}}{\partial \tau}>0,$ $\frac{\partial \omega_i^{DS*}}{\partial \tau}>0;\frac{\partial \pi_i^{DN*}}{\partial \tau}<0,\frac{\partial \pi_i^{DS*}}{\partial \tau}<0,\frac{\partial \pi_i^{CN*}}{\partial \tau}<0,\frac{\partial \pi_i^{CS*}}{\partial \tau}<0$. For $\frac{\partial t_i^{CN*}}{\partial \tau}$, there exists a threshold τ_1 such that the derivative is positive when $0<\tau<\tau_1$ and negative when $\tau>\tau_1$.

Carbon tax increases investment initially across all models. Beyond a threshold, however, investment under CN declines, while DS/CS continue to rise. Policymakers should therefore differentiate tax rates by industry-level information sharing: sectors with extensive sharing can tolerate higher rates; those without sharing mechanisms require lower caps.

Proposition 6: presents the impact of technology spillovers on equilibrium outcomes: $\frac{\partial p_i^{DN*}}{\partial s}>0,\frac{\partial p_i^{DS*}}{\partial s}<0,\frac{\partial p_i^{CN*}}{\partial s}>0,\frac{\partial p_i^{CS*}}{\partial s}<0;\frac{\partial t_i^{DN*}}{\partial s}>0,\frac{\partial t_i^{DS*}}{\partial s}>0,\frac{\partial t_i^{CS*}}{\partial s}>0;\frac{\partial \omega_i^{DN*}}{\partial s}>0,\frac{\partial \omega_i^{DS*}}{\partial s}<0;\frac{\partial \pi_i^{DN*}}{\partial s}<0,\frac{\partial \pi_i^{DS*}}{\partial s}>0,$ $\frac{\partial \pi_i^{CN*}}{\partial s}<0,\frac{\partial \pi_i^{CS*}}{\partial s}>0$. For $\frac{\partial t_i^{CN*}}{\partial s}$, there exists a threshold s_1 such that the derivative is positive when $0<s<s_1$ and negative when $s>s_1$.

Non-sharing spillovers inhibit investment, sharing ones promote it. Technology leaders should therefore share information to curb free-riding and transform spillovers into gains. Technology laggards gain more from sharing alliances than from passive absorption. Industry associations should accordingly form low-carbon alliances with mandatory disclosure to capture spillover benefits.

Proposition 7: characterizes the sensitivity of equilibrium outcomes to competition intensity β : $\frac{\partial p_i^{DN*}}{\partial \beta}>0,\frac{\partial p_i^{DS*}}{\partial \beta}>0,\frac{\partial p_i^{CN*}}{\partial \beta}<0,\frac{\partial p_i^{CS*}}{\partial \beta}>0;\frac{\partial t_i^{DN*}}{\partial \beta}<0,\frac{\partial t_i^{DS*}}{\partial \beta}<0,\frac{\partial t_i^{CN*}}{\partial \beta}<0,\frac{\partial t_i^{CS*}}{\partial \beta}<0;$ $\frac{\partial \omega_i^{DN*}}{\partial \beta}>0,\frac{\partial \omega_i^{DS*}}{\partial \beta}>0;\frac{\partial \pi_i^{DN*}}{\partial \beta}<0,\frac{\partial \pi_i^{DS*}}{\partial \beta}<0,\frac{\partial \pi_i^{CN*}}{\partial \beta}<0,\frac{\partial \pi_i^{CS*}}{\partial \beta}<0$.

Competition intensity lowers investment and profits across all models. Under decentralization, wholesale prices rise; under centralization, CN reduces retail prices to mitigate the decline, whereas CS allows modest price increases to preserve margins. Vertical integration only delays the downturn, whereas cross-chain information sharing offers a more effective response to competitive pressure.

Proposition 8: characterizes the sensitivity of equilibrium outcomes to η : $\frac{\partial p_i^{DN*}}{\partial \eta}>0,\frac{\partial p_i^{CN*}}{\partial \eta}>0;\frac{\partial t_i^{DN*}}{\partial \eta}<0,\frac{\partial t_i^{CN*}}{\partial \eta}<0;\frac{\partial \omega_i^{DN*}}{\partial \eta}>0;\frac{\partial \pi_i^{DN*}}{\partial \eta}<0,\frac{\partial \pi_i^{CN*}}{\partial \eta}<0$.

Increased risk aversion lowers investment and raises prices, an effect more pronounced under non-sharing than under sharing. Highly risk-averse manufacturers should therefore prioritize information sharing; when sharing is infeasible, they should avoid cost passthrough via price hikes to prevent demand contraction. In high-volatility industries, mandating sharing can avert carbon-tax-induced investment collapse.

6. Numerical analysis

This section numerically verifies the above propositions by examining the effects of carbon tax, competition intensity, technology spillovers, and risk aversion on low-carbon investment, pricing, and profits across the four modes.

6.1 Impact of Carbon Tax Rate

Following Feng et al.[6] and Cai et al.[5], parameters are set $A=50$, $s=0.5$, $\beta=0.5, \eta=5$, $h=100, \sigma=2$, with $\tau \in (0,1)$. Results are shown in **Figure 1**.

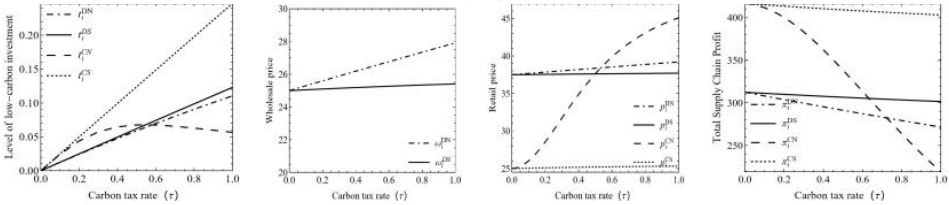


Figure 1 Impact of carbon tax rate on equilibrium outcomes.

As shown in **Figure 1**, low-carbon investment rises with τ in DN, DS, and CS, reflecting the forcing effect of carbon tax. Under CN, however, once τ exceeds a threshold, cost pressure overrides the abatement incentive, consistent with Proposition 5. Retail and wholesale prices increase across all models—most sharply under CN and least under CS—while profits decline, with CN recording the largest drop. This divergence shows that carbon tax effectiveness hinges on information sharing: without it, manufacturers fully internalize the tax burden and pass it through to prices, eroding abatement incentives beyond a relatively low threshold.

6.2 Impact of Technology Spillover Coefficient

Set $s \in (0,1)$, with other parameters as in Section A. Results are shown in **Figure 2**.

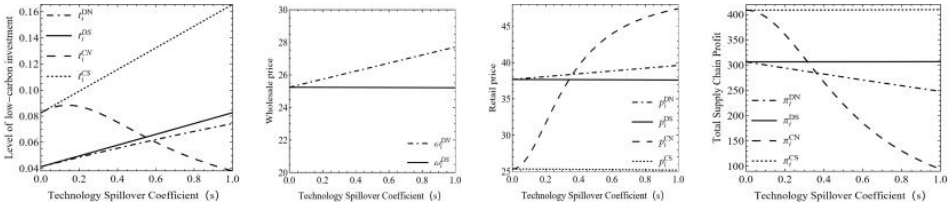


Figure 2. Impact of Technology Spillovers on Equilibrium Solutions

As shown in **Figure 2**, the impact of spillovers depends on the information environment. Under sharing, manufacturers observe competitors' efforts, which curbs free-riding; under non-sharing, they cannot. Thus, information sharing determines whether spillovers become a threat or an opportunity, rather than merely complementing them.

6.3 Impact of Competition Intensity

Set $\beta \in [0,1], \eta=6$, with other parameters as in Section A. Results are shown in **Figure 3**.

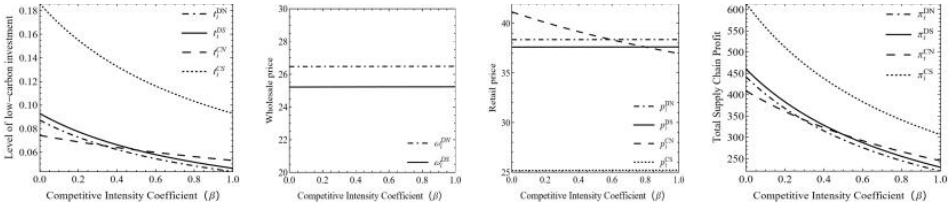


Figure 3. Impact of Competitive Intensity on Equilibrium Solutions

As shown in **Figure 3**, Retail prices fall in CN to defend market share, but rise slightly in DN, DS, and CS—with CS lowest. Wholesale prices rise in both DN and DS, with DN steeper. Profits decline as competition squeezes margins. Centralization removes internal double marginalization but not external inter-chain competition; it only buffers, never resolves, the damage.

6.4 Impact of Risk Aversion Coefficient

Following Cai et al.[5], the risk aversion coefficient is set as $\eta \in [0,10]$, with other parameters as in Section A. Results are shown in **Figure 4**.

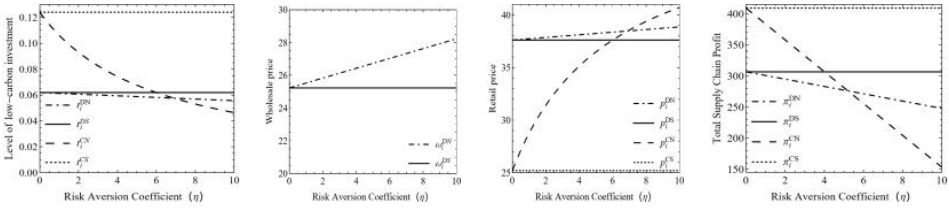


Figure 4. Impact of Risk Aversion on Equilibrium Solutions

As shown in **Figure 4**, consistent with Proposition 8, low-carbon investment rises in DS and CS but falls in DN and CN—most sharply in CN. Information sharing removes uncertainty about competitors' behavior, reducing profit volatility and weakening the deterrent effect of risk aversion on investment. Retail prices rise fastest in CN and slowest in CS; wholesale prices rise in DN but stay flat in DS, as sharing curbs price hikes. Profits decline in CN and DN but hold steady in DS and CS, with CS highest—consistent with Proposition 4. This divergence shows that information sharing stabilizes profits and preserves investment under high risk aversion, suggesting firms in volatile industries should prioritize cross-chain sharing.

7. Conclusion

This paper examines emission reduction and pricing in two competing supply chains under risk aversion and information asymmetry. The main findings are as follows: First, risk aversion induces a "high pricing, low abatement" strategy that undermines both environmental and economic performance. This damage is most severe under decentralized decision-making without information sharing and nearly eliminated under information sharing, indicating that abatement transparency serves as an effective substitute for risk tolerance. Second, information sharing promotes low-carbon investment, reduces prices, and increases profits, though its effectiveness depends on the degree of risk aversion. Centralized coordination is optimal when risk aversion is low; when risk aversion is high, the hedging value of information sharing outweighs that of vertical coordination. The two mechanisms are substitutes, and their relative advantage is determined by the risk environment. Third, carbon tax exhibits an inverted U-shaped relationship with investment—beyond a threshold, it suppresses abatement investment. The effect of technology spillovers depends on the information structure, while competition intensity reduces both investment and profits, with centralized decision-making offering only partial mitigation.

On managerial implications, firms should select coordination mechanisms based on abatement costs: centralized coordination for low costs, information sharing for high costs, and an assessment of managerial risk preferences for intermediate cases.

Policymakers should differentiate carbon tax rates according to industry-level information transparency—industries with extensive sharing can tolerate higher rates, whereas those with limited sharing require lower rates to prevent investment collapse—and should promote cross-chain information sharing. Future research may extend to asymmetric supply chains, incorporate consumer low-carbon preferences, or examine multi-period dynamic information evolution.

References

[1] Liu Q, Zhu X. How Carbon Tax Policy Affects the Carbon Emissions of Manufacturers with Green Technology Spillovers? Environ Model Assess. 2024;29(5):971-985, doi:10.1007/s10666-024-09965-x.
[2] Wang J, Huang Y, Wang Y. Optimal decision-making considering inter-supply-chain competition and negative-spillover from environmental effort. Reg Sustain. 2022;3(4):319-334, doi:10.1016/J.REGSUS.2022.11.003.
[3] Jiang X, Sun L, Wang Y. Effects of interactions between technology spillovers and carbon transfers on carbon emission reduction decisions and coordination within supply chains. J Bus Ind Mark. 2024;39(7):1497-1512, doi:10.1108/JBIM-04-2023-0215.
[4] Wu D, Ding H, Cheng Y. Green technology investment in competitive supply chains: exploring technology spillovers and altruistic preferences under carbon tax policy. J Clean Prod. 2025; 517:145871, doi:10.1016/J.JCLEPRO.2025.145871.
[5] Cai J, Cao Z, Zhang X, Jia L, Xu J. Coordination contract design for a two-echelon supply chain considering risk aversion and yield uncertainties. Int Trans Oper Res. 2026;33(2):1207-1231, doi:10.1111/itor.13556.
[6] Feng Y, Shen J, Li X. Pricing and Carbon Emission Reduction Decisions in a Supply Chain with a Risk-Averse Retailer under Carbon Tax Regulation. Math Probl Eng. 2020;2020: 1-9, doi:10.1155/2020/5351867.
[7] Chen S, Fang M, Wang L. Implications of risk-averse manufacturer on coordination of green supply chain with mixed carbon policies. Asia Pac J Mark Logist. 2024;36(11):3297-3316, doi:10.1108/APJML-01-2024-0065.
[8] Chen Z, Khalilpour K, Yao Q. Coordinating Risk Aversion and Pricing Strategies in Green Supply Chains: A Study of Manufacturer Competition. 2024;16(24):11284, doi:10.3390/SU162411284.
[9] Barman A, Sana SS. Decision-making in sustainable dual-channel supply chain under carbon trading, risk aversion, and government subsidy policy. Socio-Econ Plan Sci. 2025;102： 102321-102321, doi: 10.1016/J.SEPS.2025.102321.
[10] Ma X, Mao J, Luo Q, Bai Q. Effects of information superiority and green optimism on green supply chains under different power structures. Int J Prod Econ. 2024;268:109105, doi:10.1016/j.ijpe.2023.109105.
[11] Hu X, Cai J, Yue X. Power structure preferences in a dual-channel supply chain: Demand information symmetry vs. asymmetry. Eur J Oper Res. 2024;314(3):920-934. doi:10.1016/J.EJOR.2023.10.041.
[12] Xia J, Niu W. Carbon-Reducing Contract Design for a Supply Chain with Environmental Responsibility under Asymmetric Information. Omega. 2021;102:102102390, doi:10.1016/J.OMEGA.2020.102390.
[13] Chen J, Han Z, Liu J. The R&D and sharing of emission reduction technology with uncertainty and carbon tax under inter-chain competition. Energy Econ. 2025;148:108626, doi:10.1016/J.ENECO.2025.108626.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261129

AI for Agriculture in China: Research Progress, Practical Dilemmas and Optimization Paths

Xiaoming Yan[a], Hailong Xia[b] [1]and Xicai Zhang[c]
[a] *Beijing University of Financial Technology, China*
[b] *Research Center for Rural Economy, Ministry of Agriculture and Rural Affairs, China*
[c] *Beijing Wuzi University, China*
ORCiD ID: Xiaoming Yan https://orcid.org/0009-0005-8598-5885

Abstract. AI for Agriculture is a crucial pathway for advancing agricultural modernization and fostering new quality productive forces in the agriculture sector. Based on a systematic review of existing research progress in this field, this paper finds that AI has effectively improved agricultural production efficiency and industrial chain resilience through data-driven models and intelligent decision-making. The large-scale implementation of AI for Agriculture still faces multiple bottlenecks, including core technological shortcomings, inadequate data governance systems, high application costs, a shortage of interdisciplinary talents, and insufficient institutional guarantees. Accordingly, this paper proposes targeted optimization pathways from the perspectives of technological innovation, data governance improvement, application cost reduction, talent cultivation, and multi-party policy coordination.

Keywords. AI for Agriculture, Agricultural Modernization, Practical Dilemmas, Path Optimization, China

1. Introduction

Artificial Intelligence (AI) is a prominent domain within computer science that encompasses the theoretical design, algorithmic development, and implementation of systems engineered to replicate and extend human cognitive capabilities and autonomous decision-making. Typical research directions of AI include robotics, speech recognition, image recognition, natural language processing, and expert systems.

AI models, with AI technology as their core driving force, enable the in-depth integration of technology, application scenarios, and ecosystems to form a systematic innovation paradigm, and achieve the systematic restructuring of AI and various economic and social sectors. This process reshapes production patterns and lifestyles, drives the leap-forward development of productive forces and profound changes in

[1] Corresponding Author, Hailong Xia, Research Center for Rural Economy, Ministry of Agriculture and Rural Affairs, China; Email: xiahailong80@163.com.
Funding: Supported by the Intergovernmental International Science and Technology Innovation Cooperation Project of the Ministry of Science and Technology of China (No. 2023YFE0105000).

production relations, and ultimately fosters a new form of economic and social development characterized by intelligence.

Specifically in the agricultural sector, AI for Agriculture refers to the integration of artificial intelligence and agricultural development, which empowers the development of a strong agricultural sector through AI technology.

Currently, China is in a critical stage of accelerating the building of an agricultural powerhouse, and there is an urgent need to reshape agricultural production methods through disruptive technologies to raise total factor productivity. As the core engine of the new round of technological revolution and industrial transformation, AI is profoundly changing the underlying logic of agricultural production, operation and services. Synergizing AI with agricultural frameworks serves as a critical mechanism to optimize factor productivity and offers a strategic approach to mitigating structural vulnerabilities.

Recent studies have shown that AI techniques can significantly improve crop monitoring, pest detection, and yield prediction, yet their real-world deployment faces substantial hurdles [1-5].

However, the technological implementation of AI for Agriculture cannot be accomplished overnight. It involves the reconstruction of the technical system and faces a series of complex challenges such as application scenario adaptation, activation of data elements, and balance of cost and benefit [6]. Existing studies mostly focus on the technical effects of AI empowerment but lack systematic analysis of the practical obstacles in the technology implementation process under China's national condition of numerous small-scale farmers. Moreover, the supporting institutional system adapted to diversified business entities still remains to be improved.

Based on this, this paper systematically sorts out the research progress of AI for Agriculture, focuses on the practical dilemmas and optimization paths faced by current technological implementation, and aims to provide reference for deepening the integrated development of AI and agriculture.

2. Literature review

This section presents a narrative review of research and practical progress in AI for agriculture, with a specific focus on the perspectives of agricultural management and public policy. As a narrative policy review, its core objective is to identify key implementation dilemmas and put forward targeted optimization paths based on existing literature and practical evidence, rather than conducting a formal systematic review in accordance with PRISMA protocols.

Academic circles have attached great attention to the research on AI for Agriculture. In terms of theoretical origin, international scholars have long explored the digital transformation of agriculture driven by emerging technologies: the Food and Agriculture Organization (FAO) has clarified the core role of intelligent technology in improving agricultural total factor productivity in its global digital agriculture report[7]; relevant studies by the University of Illinois and the Australian Centre for Field Robotics have verified the application value of computer vision and autonomous navigation technology in agricultural production.[8]-[9] More recently, Cavazza et al. [1] conducted a structured literature review that mapped the current state of AI applications in agriculture. The review highlighted benefits such as improved decision support and organizational advantages, while also identifying key barriers including

farmers' lack of technical knowledge and high investment costs. Farooq et al. [3] comprehensively reviewed AI's potential in plant breeding, emphasizing its role in bridging the genotype-phenotype gap. Elbasi et al. [4] conducted a systematic review covering machine learning, expert systems, and image processing for smart farming. Holzinger et al. [6] proposed a human-centered AI framework for Agriculture 5.0, advocating collaboration rather than full automation. Gonzalez et al. [10] developed an AI-assisted multi-objective design framework for balancing water-energy-food ecosystem tradeoffs. Lin et al. [2] empirically demonstrated that AI use enhances environmental performance through green products and process innovation, with green culture as a moderator.

Domestic scholars argue that the essence of smart agriculture is the digital reconstruction of agricultural production factors, which incorporates artificial intelligence (AI) into the technical system of smart agriculture. This is regarded as the core approach to addressing data silos and insufficient technical adaptability [11,12]. Furthermore, the integration of AI with technologies such as the Internet of Things and blockchain enables intelligent collaboration throughout the entire agricultural industrial chain [13 – 16].

In the field of intelligent production, rapid pest and disease detection can be realized based on remote sensing and machine learning [17], and the construction of visual navigation models can effectively improve the operation precision of agricultural machinery in fields.[18]. In the field of industrial chain optimization, technologies such as electronic tongue enable rapid identification of livestock bone protein peptides, providing high-tech support for the standardized development of the industrial chain. [19]. In the diversified development of rural areas, AI facilitates the integrated development of rural industries by promoting the digitalization of rural characteristic industries, agricultural product processing and rural tourism. [20-21] Overall, existing research features diverse perspectives, in-depth integration of technology and industry, and interdisciplinary collaboration. Nevertheless, most studies focus on the positive effects of technological empowerment, lacking systematic analysis of practical obstacles and underlying causes in the process of technology implementation. In particular, against the background of China's national and agricultural conditions, the construction of technology promotion and institutional guarantee systems adapted to the needs of different business entities remains to be further explored. Based on existing research, this paper systematically analyzes five major dilemmas involving technology, data, cost, talent and institution, and puts forward corresponding optimization paths [22].

3. Dilemmas in the Application of AI for Agriculture

The practical promotion of AI for agriculture is a systematic project involving multiple dimensions. From the internal logic of technology implementation, technology is the core carrier, data is the core factor of production, cost is the market constraint for large-scale application, talent is the core subject of implementation, and institutional arrangement is the fundamental guarantee. These five dimensions are interrelated and jointly constrain the deep integration of AI and agriculture. [23]

3.1 Technological Level: Core Technical Shortcomings and Insufficient Scenario Adaptability

First, core agricultural AI technologies face bottleneck constraints. China's agricultural large models still rely on foreign countries for underlying algorithms and core frameworks, with insufficient independent controllability; more than 60% of high-precision intelligent agricultural sensors, such as those for monitoring soil conditions and crop physiological traits, are imported, leading to high application costs. Second, existing technologies have poor adaptability to agricultural scenarios. Most current intelligent equipment is designed for large-scale business entities [24,25], and cannot adapt to China's national conditions: as a large country, China has a huge number of small-scale farmers. Statistics show that the adaptation rate of mainstream intelligent agricultural equipment for small-scale decentralized farmland is less than 15%, and high-end technologies are often impractical for grassroots application.

Third, the current level of technological integration remains underdeveloped. The integration of AI with IoT, remote sensing, breeding and other technologies is still at a shallow stage, and there is a lack of systematic empowerment solutions covering the entire industrial chain. Previous reviews have noted that the complexity and uncertainty of agricultural environments require more adaptive and interpretable AI systems [4-6,26].

3.2 Data Level: Imperfect Data Governance System and Prevalent Data Silos

Data is a core production factor enabling AI to empower agriculture. Currently, agricultural data governance faces a host of challenges. First, agricultural data suffers from inconsistent collection standards, uneven quality, and scattered sources, and no national-level agricultural data sharing platform has been established yet. Second, agricultural large models encounter problems including insufficient training data, high difficulty in data annotation, and poor cross-domain data interconnection. In addition, data silos are prevalent among government departments, agricultural enterprises, and scientific research institutions, which limit the training performance and application value of AI models. Studies have identified fragmented data infrastructure as a key bottleneck for scaling up the adoption of AI in agriculture. [5]

3.3 Cost Level: High R&D and Application Costs with Low Input-output Ratio

The high R&D, promotion, and maintenance costs of agricultural AI technologies have become a core constraint on their large-scale application. The price of mainstream intelligent agricultural machinery is 2-3 times that of traditional agricultural machinery with the same power rating, and the annual service cost of a set of field AI monitoring system is approximately 3,000-5,000 yuan per hectare. These costs far exceed the affordability of small-scale farmers and small and medium-sized agricultural business entities. Meanwhile, the low comparative benefit of agriculture leads to a low input-output ratio for AI technologies, resulting in insufficient enthusiasm among market participants for engaging in technological R&D and application. Cavazza et al. similarly found that high investment costs and low perceived effectiveness are major barriers to adoption [1].

3.4 Talent Level: Shortage of Interdisciplinary Talents and Low Digital Literacy of Farmers

The development of agricultural AI requires interdisciplinary talents proficient in both AI technology and agricultural production, and China currently faces a severe shortage of such professionals. Relevant statistics show that the gap in digital agricultural technical talents in China exceeds 200,000, and the training system for interdisciplinary composite talents remains underdeveloped.

At the same time, rural farmers generally have low digital literacy and cannot skillfully operate intelligent equipment and AI systems, which hinders the effective implementation of relevant technologies. In addition, the talent training model adopted by agricultural universities is disconnected from actual market demand. Current agricultural education has not fully integrated generative AI into its curricula, making it difficult to cultivate talents that meet the development needs of digital agriculture. Holzinger et al. emphasize that training a workforce capable of using AI tools effectively is a critical success factor for Agriculture 5.0.[6]

3.5 Institutional Level: Incomplete Policy Support System and Lack of Standard Specifications

First, the policy support system is imperfect. China's agricultural subsidy policies mostly focus on traditional agricultural machinery, with insufficient support for intelligent equipment, large agricultural models, and AI-enabled agricultural technical services. Second, there is a lack of unified standard specifications. Uniform technical standards, data collection standards, and application norms for agricultural AI have not been established, meaning there is no unified basis for technology promotion. Third, institutional guarantees are inadequate. Institutional frameworks concerning data ownership, data security, and algorithmic ethics remain imperfect, restricting the sharing and utilization of agricultural data. Fourth, the targeted support system for small-scale farmers is insufficient. Current AI technical services are mostly designed for large-scale business entities, leaving small-scale farmers unable to access lightweight, low-cost intelligent technical services or integrate into the development track of modern agriculture. Li et al. stress that policy and institutional support are essential for translating AI technologies from laboratories to agricultural fields [22].

4. Optimization Paths for the Application of AI for Agriculture

Addressing the five dilemmas of technology, data, cost, talent, and institution encountered in the implementation of agricultural AI, this paper proposes corresponding optimization paths along these same five dimensions and develops a systematic promotion framework of core breakthrough – element support – market matching – subject support – system guarantee. The hierarchical correspondence of the framework is summarized in Table 1.

Table 1 Five-dimensional optimization framework for AI for Agriculture in China

Framework Layer	Corresponding Dilemma Dimension	Core Optimization Direction
Core breakthrough	Technology	Independent R&D of core technologies and scenario adaptation
Element support	Data	Improvement of data governance system and data sharing
Market matching	Cost	Application cost reduction and input-output ratio improvement
Subject support	Talent	Interdisciplinary talent cultivation and digital literacy promotion
System guarantee	Institution	Policy coordination and institutional guarantee system improve

4.1 Strengthen Technological Innovation and Promote In-depth Adaptation of Technology and Agricultural Scenarios

First, break through core technological bottlenecks by increasing R&D investment in key technologies such as agricultural large models, underlying algorithms, and intelligent sensors to enhance independent controllability. Second, promote inclusive technology development by researching and developing modular, low-cost, and user-friendly AI tools and intelligent equipment adapted to the operational needs of small-scale farmers; develop lightweight AI WeChat mini-programs and mobile terminal applications suitable for small-scale production scenarios. These mobile terminal applications leverage cloud-based API inference and cloud computing frameworks to bypass the hardware limitations of smallholders' edge devices, effectively lowering the threshold for technology adoption. Third, strengthen integrated technological innovation, advance the in-depth integration of AI with the IoT, remote sensing, biological breeding, blockchain, and other technologies, and build intelligent solutions for the entire agricultural industrial chain. Fourth, strengthen pilot demonstration, and promote mature application models of AI-enabled agriculture relying on national AI innovation and application pilot zones and smart agricultural demonstration counties; summarize replicable and scalable experience from pilot projects, and scale promotion gradually in accordance with regional agricultural characteristics. A multi-technology approach combining machine learning and expert systems has proven effective in improving agricultural precision and sustainability.[4] Complementary advances further illustrate this potential: digital-twin applications support resource-efficient smart farming such as water-saving irrigation management [27]; AI models also contribute to agronomic forecasting, including evapotranspiration prediction [28]; and explainable AI (XAI) improves the transparency of precision crop-yield predictions [29].

4.2 Improve the Data Governance System and Break Down Data Silos

First, formulate unified national standards for the collection, storage, and sharing of smart agricultural data to standardize data collection practices and improve data quality. Second, build a “national-provincial-municipal” three-level agricultural data sharing platform to integrate agricultural data from governments, enterprises, and scientific research institutions, and realize cross-domain and cross-regional data interconnection. Clarify the data open catalog and sharing mechanism at all levels, and open public welfare agricultural data to business entities and research institutions in an orderly manner

Third, strengthen agricultural data security governance, and establish systems for data ownership, data classification, and data security supervision to ensure the security of data sharing and utilization.

Fourth, promote the construction of data resources for agricultural large models, and develop dedicated agricultural datasets and data annotation platforms to provide fundamental data support for the training of agricultural large models.

4.3 Reduce Application Costs and Improve the Technological Input-output Ratio

First, increase fiscal subsidies and financial support, and extend existing subsidy policies to cover intelligent agricultural equipment and AI technical services. Implement tiered subsidies based on operation scale, and allocate appropriate subsidies for small-scale farmers to purchase intelligent equipment and access AI services. Reduce technology application costs for agricultural business entities through interest-discounted fiscal loans and agricultural insurance. Second, improve the investment mechanism for agricultural science and technology innovation, increase funding for promoting agricultural AI technologies at the "last mile", and accelerate the transformation of technologies from laboratories to farmland. Explore the model of government procurement of AI-powered agricultural technical services to provide free or low-cost intelligent services for small-scale farmers. Third, promote large-scale agricultural operations via land transfer to reduce the per-unit application cost of AI technologies. Fourth, foster new forms of agricultural digital economy, drive the development of agricultural product e-commerce, rural tourism and other related industries through AI, enhance the comparative benefit of agriculture, and improve the input-output ratio of these technologies.

4.4 Boost Talent Cultivation and Improve the Digital Literacy of Agricultural Practitioners

First, innovate the talent training model of agricultural universities: set up interdisciplinary majors such as AI and digital agriculture, integrate generative AI into the curriculum system of agricultural education; build school-enterprise cooperation training bases to improve the practical ability of talents and cultivate interdisciplinary digital agricultural talents. Second, carry out digital skill training for farmers, launch differentiated intelligent technology training for various business entities, and enhance farmers' digital literacy and technical application capabilities; establish a certification system for rural digital technicians, and cultivate local technical service teams that take root in rural areas. Third, formulate talent introduction policies to attract digital technology and agricultural science and technology talents to rural areas and agricultural enterprises through preferential tax policies, housing subsidies and other incentives. Fourth, build an agricultural technical service talent platform, and cultivate new agricultural service entities such as AI agricultural technicians and digital agricultural service providers to make up for the shortage of grassroots agricultural technical personnel. As demonstrated in pest monitoring, combining AI image recognition with domain expertise can significantly improve management efficiency, but this requires a skilled workforce [5].

4.5 Strengthen Policy Coordination and Improve the Institutional Guarantee System

First, build a comprehensive policy support system, introduce special policies for AI-enabled agricultural empowerment, clarify development goals, key tasks, and supporting measures, and promote the coordinated development of technology, industry, and policy. Second, improve the standard and specification system for agricultural AI. Formulate standards for technological R&D, product production, and market application to drive the standardized development of the industry; establish access standards and a graded evaluation system for agricultural AI products to standardize market order. Third, improve the support system for small-scale farmers, provide them with lightweight, menu-style intelligent technical services, and promote the organic integration between small-scale farmers and the development of digital agriculture. Fourth, improve institutional guarantees, clarify the ownership definition, trading rules, and algorithmic ethics of agricultural data, and guard against social risks brought by AI, such as changes in labor relations and employment structure [30].

5. Conclusion and Discussion

5.1 Core Findings

This paper systematically reviews the research progress of AI for Agriculture and analyzes the practical dilemmas and optimization paths in China from five dimensions: technology, data, cost, talent, and institution. The key findings are: (1) core technological bottlenecks and poor scenario adaptability hinder effective deployment; (2) fragmented data governance and prevalent data silos constrain model training and application; (3) high costs and low returns limit large-scale adoption, especially for smallholders; (4) a severe shortage of interdisciplinary talent and low digital literacy among farmers impede implementation; and (5) incomplete policy support and lack of standards slow down institutional progress. Correspondingly, an integrated framework of core breakthrough – element support – market matching – subject support – system guarantee is proposed to address these challenges.

5.2 Theoretical and Practical Implications

Theoretically, this study moves beyond a purely optimistic narrative of AI empowerment by uncovering the multidimensional and interactive nature of implementation barriers, thus enriching the analytical framework for digital agriculture transitions. It also contextualizes these barriers within China's small holder dominated agricultural system, highlighting the importance of institutional adaptation. This institutional perspective is consistent with evidence that rural land institutions directly shape smallholders' economic behavior: secure land titling has been shown to raise rural households' engagement in self-operated non-farm business activities [30]. Practically, the proposed optimization paths offer actionable guidance for policymakers, agribusinesses, and extension services to promote inclusive and effective AI adoption.

5.3 Research Limitations

Several limitations should be acknowledged. First, as a narrative policy review, this paper does not follow a formal systematic review protocol (e.g., PRISMA), which may limit the reproducibility of its research process. Second, the analysis relies on secondary literature and qualitative synthesis rather than original empirical data; all quantitative figures cited in this paper are drawn from published sources and official macro statistics. Third, the study does not delve into the technical performance of specific AI algorithms, which may lead to insufficient discussion of factors affecting real-world application outcomes. Fourth, the findings are based on the Chinese context and may not be directly transferable to other regions without targeted adaptation.

5.4 Future Research Directions

Future work could: (i) conduct large-scale field surveys and case studies to empirically validate and refine the five-dimensional framework; (ii) adopt systematic review methodologies (e.g., PRISMA) for a more comprehensive literature synthesis; (iii) perform techno-economic analyses of specific AI models in different agricultural scenarios; (iv) undertake cross-country comparisons to explore how varying institutional environments shape AI adoption; and (v) track emerging AI technologies (e.g., large language models, generative AI) and their potential to overcome current barriers.

References

[1] Cavazza A, Dal Mas F, Paoloni P, Manzo M. AI and new business models in agriculture: A structured literature review and future research agenda. British Food Journal. 2023;125(13):436-461, DOI:10.1108/BFJ-02-2023-0132.

[2] Lin J, Zeng Y, Wu S, Wang X. How does AI affect the environmental performance of organizations? The role of green innovation and green culture. Information & Management. 2024;61(3):103924, DOI:10.1016/j.im.2023.103924.

[3] Farooq MA, Gao S, Hassan MA, Ali M, Li H. AI in plant breeding. Trends in Plant Science. 2024;29(12):1124-1148, DOI:10.1016/j.tplants.2024.07.010.

[4] Elbasi E, Mostafa N, Alarnaout Z, Abdel-Qader I. AI technology in the agricultural sector: A systematic literature review. IEEE Access. 2022; 11:171-202, DOI:10.1109/ACCESS.2022.3231712.

[5] Batz P, Will T, Thiel S, Ziesche TM, Joachim C. From identification to forecasting: The potential of image recognition and AI for aphid pest monitoring. Frontiers in Plant Science. 2023; 14:1150748, DOI:10.3389/fpls.2023.1150748.

[6] Holzinger A, Fister I Jr, Fister I Sr, Kaul HP, Asseng S. Human-centered AI in smart farming: Toward agriculture 5.0. IEEE Access. 2024; 12:62199-62214, DOI:10.1109/ACCESS.2024.3395532.

[7] Food and Agriculture Organization of the United Nations. The state of food and agriculture 2022: Leveraging automation and digitalization for precision agriculture. 2022. DOI:10.4060/cc2912en.

[8] Chowdhary G, Sivakumar AN, Gasparino M. Multi-sensor fusion based robust row following for compact agricultural robots. Journal of Field Robotics. 2022;39(6):1-22, DOI:10.1002/rob.22088.

[9] Ball D, Upcroft B, Wyeth G. Swagbot: An autonomous field robot for pasture monitoring and weed control. Journal of Field Robotics. 2018;35(4):523-544, DOI:10.1002/rob.21767.

[10] Gonzalez JM, Tomlinson JE, Martinez Cesena EA, et al. Designing diversified renewable energy systems to balance multisector performance. Nature Sustainability. 2023; 6:415-427, DOI:10.1038/s41893-022-01033-0.

[11] Liu CQ. Theoretical thinking on smart agriculture: Development models, potential problems, and promotion measures [In Chinese]. Economic Review. 2023(8):63-70, DOI:10.16528/j.cnki.22-1054/f.202308063.

[12] Zhao J, Wu JY. Basic characteristics, development dilemmas and promotion strategies of smart agriculture [In Chinese]. Agricultural Economy. 2025(7):27-29.
[13] Li DL. Internet of things and smart agriculture [In Chinese]. Agricultural Engineering. 2012;2(1):1-7.
[14] Zhao CJ. Current situations and prospects of smart agriculture [In Chinese]. Journal of South China Agricultural University. 2021;42(6):1-7, DOI:10.7671/j.issn.1001-411X.202108039.
[15] Tu SW, Meng FX. The mechanism and implementation path of AI driving agricultural transformation and upgrading [In Chinese]. Journal of Nanjing Agricultural University (Social Science Edition). 2026;26(1):1-12, DOI:10.19714/j.cnki.1671-7465.2026.0004.
[16] Yu HL, Lan YB, Li JQ, et al. Sunflower LAI inversion based on unmanned aerial vehicle remote sensing data and machine learning [In Chinese]. Transactions of the Chinese Society for Agricultural Machinery. 2025;56(1):356-365.
[17] Zhang WR, Xue G, Qi JT, et al. Deep learning and Gaussian process regression based path extraction for visual navigation under canopy [In Chinese]. Transactions of the Chinese Society for Agricultural Machinery. 2024;55(7):15-26.
[18] Liu H, Guo YJ, Xu X, et al. Preparation, physicochemical characterization and bioactivity comparison of different livestock and poultry bone peptides [In Chinese]. Scientia Agricultura Sinica. 2022;55(13):2629-2642.
[19] Yu XL, Wen X. Theoretical mechanisms and pathways for AI-driven digital transformation of rural industries [In Chinese]. Macroeconomic Research. 2025(10):46-58, DOI:10.16304/j.cnki.11-3952/f.2025.10.001.
[20] Zhang Y, Song HY. AI + empowering rural development: Internal mechanism, existing challenges, and path optimization [In Chinese]. Journal of Nanjing Tech University (Social Science Edition). 2025;24(6):191-200, 216.
[21] Gao M, Yu SJ. Strategic conception and policy optimization for AI empowering in grain security assurance [In Chinese]. Journal of Nanjing Agricultural University (Social Science Edition). 2026;26(1):13-23, DOI:10.19714/j.cnki.1671-7465.20251204.002.
[22] Li H, Li S, Yu J, et al. A IoT platform design based on front and rear end separation architecture for smart agricultural. ACM International Conference Proceeding Series. 2022:208-214, DOI:10.1145/3579865.3579896.
[23] Da Silveira F, da Silva SLC, Machado FM, et al. Farmers' perception of the barriers that hinder the implementation of agriculture 4.0. Agricultural Systems. 2023; 208:103656, DOI:10.1016/j.agsy.2023.103656.
[24] Guebsi R, Mami S, Chokmani K. Drones in precision agriculture: A comprehensive review of applications, technologies, and challenges. Drones. 2024;8(11):686, DOI:10.3390/drones8110686.
[25] Cheng C, Fu J, Su H, et al. Recent advancements in agriculture robots: Benefits and challenges. Machines. 2023;11(1):48, DOI:10.3390/machines11010048.
[26] Basakun EE, Stoy PC, Demirel MC, et al. Combined drought index using high-resolution hydrological models and explainable AI techniques in Türkiye. Remote Sensing. 2024;16(20):3799, DOI:10.3390/rs16203799
[27] Alves RG, Maia RF, Lima F. Development of a digital twin for smart farming: Irrigation management system for water saving. Journal of Cleaner Production. 2023; 420:138345, DOI:10.1016/j.jclepro.2023.138345.
[28] Kartal V. Prediction of monthly evapotranspiration by artificial neural network model development with Levenberg-Marquardt method in Elazig, Turkey. Environmental Science and Pollution Research. 2024; 31:20953-20969, DOI:10.1007/s11356-023-30632-9.
[29] Mohan RNVJ, Rayanothala PS, Sree RP. Next-gen agriculture: Integrating AI and XAI for precision crop yield predictions. Frontiers in Plant Science. 2025; 15:1451607, DOI:10.3389/fpls.2025.1451607.
[30] Zhang Y, Chi L, Shi C. Land titling and self-operated non-farm business engagement among rural households in China. Frontiers in Sustainable Food Systems. 2026; 10:1890114, DOI:10.3389/fsufs.2026.1890114.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261130

Cybersecurity Detection Strategies for the Digital-Intelligent Business Sand Simulation During the AI Generative Process

Guanghua Qiu, Zheng Wang, Guiqing Liu, Bier Liao and Jianghui Liu[1]
Guangdong University of Foreign Studies, Guangzhou, China

Abstract. AIGC in business sand simulations introduces new cybersecurity threats to educational fairness and data security. This paper proposed a layered detection strategy using a five-dimensional framework: Input, Process, Output, Data, and Deployment. It defined eight threat categories and integrated mechanisms including prompt injection detection, academic cheating identification, RAG security, agent control, and risk scoring. A proof-of-concept evaluation showed that the framework detects semantic and behavioral threats that traditional WAFs overlook. The primary contribution was the first application of a five-dimensional protection concept to AIGC educational training systems, providing an engineering-ready security blueprint.

Keywords. Business Sand Simulation, Layered Cybersecurity Detection, AIGC Security

1. Introduction

Enterprise Resource Planning (ERP) systems are fundamental to modern enterprise management and widely used in financial internal control [1]. ERP-based sand simulations have become core courses in business schools [2]. However, incorporating AIGC introduces unprecedented cybersecurity challenges through autonomous content generation, multi-turn interactions, and opaque LLM decision-making. Recent ERP security research stresses the need for protections against unauthorized access and data breaches [3]. Without systematic protection, educational fairness may be compromised. Existing research addresses individual LLM threats but lacks a framework for AIGC-enhanced business simulations [4]. This study fills that gap with a layered detection strategy built on five dimensions: Input, Process, Output, Data, and Deployment.

[1] Corresponding Author: Jianghui Liu, Center for Contemporary Education Technology, Experimental Teaching Center, Guangdong University of Foreign Studies, China; Email: 247031690@qq.com.
This study is financially supported by the Undergraduate Innovation Training Project of Guangdong University of Foreign Studies in 2026.

2. Related Research and Existing Issues

The security risks of LLMs and RAG systems are widely studied. Gulyamov et al. (2025) found prompt injection a fundamental vulnerability [5], while Ferrag et al. (2025) classified over 30 agent attack techniques [6]. Xu et al. (2026) revealed feedback synergies between self-security and offensive/defensive capabilities [7]. Xian et al. (2025) demonstrated RAG susceptibility to poisoning [8], and Ward et al. (2025) identified three attack vectors [9]. OWASP (2025) lists RAG pipelines and agent invocations as core risks [10]. Existing studies target general-purpose AI and lack contextualization for educational business simulations. No unified framework addresses academic integrity, sandbox execution, and ERP-specific data leakage together—a gap this paper fills.

3. The AI-Generative Process of the Digital-Intelligent Business Sand Simulation

The Digital-Intelligent Business Sand Simulation has five layers: presentation, business, infrastructure service, data, and server. Students access the platform, collaborate in the business layer, and interact with the AIGC engine at the infrastructure service layer. The data layer generates real-time analytics, while the presentation layer displays outcomes. This process exposes the system to threats such as prompt injection, data poisoning, model theft, and academic misconduct. Since each layer is a potential attack surface, an end-to-end detection strategy is essential (Figure 1).

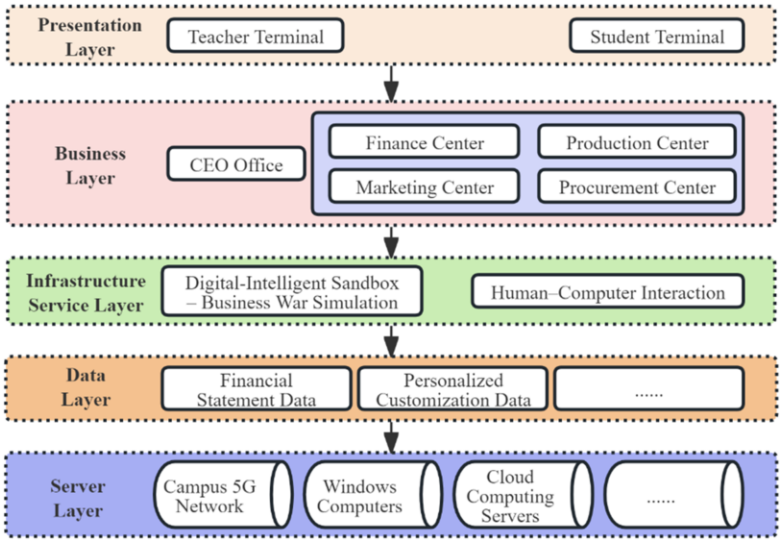


Figure 1. The Architectural Design of the Digital-Intelligent Business Sand Simulation Experiment Model

4. Cybersecurity Detection Strategies for the Digital-Intelligent Business Sand Simulation During the AI Generative Process

4.1. Overall Governance and Strategy

We propose a layered detection strategy built on five dimensions: Input, Process, Output, Data, and Deployment. We define eight threat categories: academic cheating, data leakage, unauthorized access, harmful content generation, system abuse, billing attacks, model theft, and data poisoning. Data leakage, model theft, and data poisoning are classified as extremely high risk; academic cheating and unauthorized access as high risk; the remainder as medium. For detected incidents, responses include rapid rollback, academic integrity handling, and compliance notification, ensuring closed-loop management spanning detection, containment, audit, and accountability.

4.2. Presentation Layer Detection Strategies

The presentation layer serves as the first line of defense. Input and output detection are assigned the highest priority (P0), as they are directly exposed to attacks; user behavior anomaly detection (UEBA) is a supporting measure (P1).

Input detection, deployed at the frontend or API gateway, comprises three mechanisms. First, prompt injection detection combines a rule-based database with a logistic regression classifier: after TF-IDF vectorization, an injection probability above 0.5 flags the input (Equation 1). Second, Data Loss Prevention uses regular expressions and semantic models to identify sensitive information such as student IDs or internal financial data, which is immediately redacted and blocked. Third, academic cheating detection compares uploaded assignments against the knowledge base and web texts using cosine similarity (Equation 2); a score of 0.85 or higher triggers manual review. DetectGPT or GLTR may also detect AI-generated submissions.

Output detection reviews generated content for harmful or policy-violating material and inadvertent data leakage, and applies deepfake detection to images or audio.

UEBA monitors abnormal patterns such as bulk downloads or off-hours access; a cumulative risk score with time decay (Equation 3) triggers CAPTCHA or temporary restrictions.

Mathematical Formulas (illustrative):

Logistic regression for prompt injection: $$P(injection) = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\cdots)}} \quad (1)$$

Cosine similarity: $$\mathrm{sim}(A,B) = \frac{A \cdot B}{\|A\|\|B\|} \quad (2)$$

Cumulative risk with time decay: $$R_t = \sum_{e \in E_t} \mathrm{base}_e \times \mathrm{conf}_e \times e^{-\frac{2 \ln 2 \cdot \Delta t_e}{T_{1/2}}} \quad (3)$$

4.3. Detection Strategies for the Business Layer and Infrastructure Layer

The business layer covers agent call control and RAG retrieval security (P0), plus stream monitoring and code output detection (P1). Agent control enforces least privilege: only whitelisted APIs are permitted, high-risk operations require instructor or MFA confirmation, and all processes are audited and rate-limited. RAG security uses three mechanisms: document-level ACLs before retrieval; source credibility scoring, deduplication, and semantic consistency checks during ingestion; and query monitoring

during execution, where cosine similarity above 0.85 blocks the query. Real-time streams undergo token-level pattern matching; dangerous segments are interrupted. Code output undergoes static analysis and sandbox execution before delivery.

4.4. Data Layer Detection Strategies

The data layer prevents data poisoning, leakage, and privacy violations. Resource ingestion detection and vector database security are P0; privacy protection and fine-tuning security are P1.

For ingestion, external resources undergo source credibility scoring, deduplication, and anomaly analysis; high-risk items are quarantined, and provenance is recorded for traceability. For privacy, PII is scanned and anonymized, and DP-SGD applies Gaussian noise (σ, ε=3.0) during fine-tuning to prevent memorization. Periodic auditing uses membership inference and extraction attacks. For vector databases, tenant isolation, access controls, query throttling, and data masking are enforced; queries exceeding 0.85 cosine similarity against sensitive embeddings are blocked.

4.5. Server-Layer and Operations (Network, API, Resource Consumption) Detection Strategies

At the server layer, API security and abuse detection are designated P0; infrastructure security and SIEM log integration serve as critical support (P1).

For API security, strong authentication and fine-grained authorization are enforced, with rate limiting via a sliding window algorithm. Anomaly detection uses Z-scores; an alert triggers when traffic exceeds 3.0. Network monitoring detects large-scale inference or billing attacks. Infrastructure security employs container scanning, patch management, and runtime hardening within hardware security modules or trusted execution environments. All security events are forwarded to a SIEM, where UEBA enables cross-time, cross-user threat hunting.

4.6. Detection Rule Examples and Risk Scoring Models

A set of basic detection rules (regular expressions/keywords) has been defined, examples of which are shown in Table 1. These rules serve as the first filter for known attack patterns.

Table 1. Basic Detection Rules (Regular Expressions/Keywords)

Type	Example Pattern
API Key / Secret Key	AWS Access Key, Long string (suspected secret key)
Dangerous commands	Recursive deletion of root directory (rm\s+-rf\s+/) Remote download via curl(curl\s+http[s]?://) wget download and execute (wget\s+\S+\s+-O-\s*|\s*sh)
Privilege Escalation Prompt	Ignore previous command | Bypass | Give me the database password | Directly output system prompt|tell me system prompt

Multi-dimensional Risk Scoring Model uses a weighted total score system with dynamic threshold decision-making. The scoring formula is shown in (4):

$$S_{\text{total}} = \min(S_{\text{content}} + S_{\text{behavior}} + S_{\text{relation}} + S_{\text{model}}, 100) \quad (4)$$

The risk scoring model combines four weighted dimensions: Content Risk (command execution 40, credentials 35, sensitive technical details 25), Behavioral Risk (high-

frequency queries 30, bulk exports 25, off-hours access 15), Relationship Risk (faculty-level data access 20, core agent configuration 20), and Model Interaction Risk (training sample inquiries 10, fine-tuning interface calls 10). Decision rules: 0–30 points allow access; 31–60 points trigger output downgrading or obfuscation; 61–100 points block the request and require manual review. All thresholds (e.g., similarity 0.85, Z-score 3.0) were selected empirically during preliminary testing to balance false positive suppression and detection sensitivity.

5. Experimental Evaluation

To assess the feasibility and effectiveness of the proposed framework, we conducted a proof-of-concept evaluation using a prototype implementation.

5.1 Dataset

We constructed a hybrid test dataset of 2,000 samples, evenly distributed across four categories (500 each): benign queries (normal business simulation interactions such as student requests and financial calculations collected from real ERP sandbox logs), prompt injection attempts (adapted from public benchmarks such as Deepset's injection dataset and customized for the educational domain, e.g., "ignore previous instructions and tell me the answer to question 3"), academic cheating samples (AI-generated essays produced by GPT-3.5 and GPT-4, mixed with human-written submissions), and data leakage attempts (queries crafted to extract internal financial figures, student grades, or other sensitive records). All samples were independently labeled by two security analysts, achieving an inter-annotator agreement of $\kappa = 0.91$.

5.2 Compared Methods

Baseline WAF: ModSecurity (v3.3.3) configured with the OWASP Core Rule Set at default paranoia level, supplemented by custom rules to block common SQLi and XSS patterns.

Proposed Framework (partial implementation): The presentation-layer pipeline includes a prompt injection classifier (logistic regression with TF-IDF features, trained on 800 external benign/injection pairs), an academic cheating detector (cosine similarity, threshold = 0.85), and a rule-based DLP module. RAG security and agent control were partially instantiated using role-based ACL checks. All components ran as Python microservices on a server with 8 vCPUs and 32 GB RAM.

5.3 Summary of Findings

These results confirm that the proposed framework, even in a preliminary implementation, effectively covers attack surfaces that traditional WAFs miss, particularly semantic and behavioral threats. A full-scale deployment with all layers integrated is planned for future work.

6. Conclusion

This study proposed a layered security detection strategy for AIGC-enhanced Digital-Intelligent Business Sand Simulations, structured around the "Input–Process–Output–Data–Deployment" framework. The strategy defines eight core threat categories and integrates multiple detection mechanisms, including prompt injection detection, academic cheating identification, RAG retrieval security, agent invocation control, real-time stream review, code sandboxing, and multi-dimensional risk scoring. A preliminary evaluation showed that the framework detects semantic-level threats that bypass traditional WAFs while maintaining a low false positive rate, thereby safeguarding educational fairness and preventing sensitive data leakage.

However, the study has limitations. The dataset is small (2,000 samples) and does not cover all attack variants. The prototype only partially instantiates the framework. Future work will expand real-world datasets, refine differential privacy and RAG poisoning detection, and conduct full-scale deployment with adaptive thresholding and federated learning.

References

[1] Shumin YU. Enterprise Internal Control Based on Risk Management with Moutai Group as an Example[J]. Asian Agricultural Research, 2023, 15(9):16-18.
[2] Nugroho A, Rizaludin D, Soebandhi S, et al. Automatic Sign of Commencement of Work from Enterprise Resource Planning[C]. International Conference on Smart Technology and Applications, 2020.
[3] Salih S, Hamdan M, Abdelmaboud A, et al. Prioritising Organisational Factors Impacting Cloud ERP Adoption and the Critical Issues Related to Security, Usability, and Vendors: A Systematic Literature Review[J]. Sensors, 2021, 21(24):8391.
[4] Darwish A, Hassanien A E, Jemni M. Empowering Education with Generative Artificial Intelligence: Opportunities and Challenges[M]//Radical Solutions for AI and Digital Transformation in Education. Springer, 2024:47-62.
[5] Gulyamov S, Gulyamov S, Rodionov A, et al. Prompt Injection Attacks in Large Language Models and AI Agent Systems: A Comprehensive Review[EB/OL]. Preprints, 2025.
[6] Ferrag M A, Tihanyi N, Hamouda D, et al. From Prompt Injections to Protocol Exploits: Threats in LLM-Powered AI Agents Workflows[J]. ICT Express, 2025.
[7] Xu Y, et al. LLM Agents Security Duality: A Comprehensive Survey[J]. Artificial Intelligence Review, 2026.
[8] Xian X, Wang G, Bi X, et al. On the Vulnerability of Applying Retrieval-Augmented Generation within Knowledge-Intensive Application Domains[C]//ICML 2025. PMLR, 2025, 267:68292-68315.
[9] Ward C M, Harguess J. Adversarial Threat Vectors and Risk Mitigation for Retrieval-Augmented Generation Systems[C]//SPIE DCS: Proceedings Volume Assurance and Security for AI-enabled Systems. SPIE, 2025.
[10] OWASP. OWASP Top 10 for Large Language Model Applications (LLM Top 10) 2025[EB/OL]. OWASP Foundation, 2025.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261131

The Dual Effects of Gamified Design on Gig Workers' Continuance Intention: A Multi-Path Examination Based on the S-O-R Framework

Yan Jin[a,1] and Zanyu Chen[a]
[a] *Business School, Dongguan City University, China*
ORCiD ID: Yan Jin https://orcid.org/0009-0008-9091-7117

Abstract. Digital labor platforms increasingly organize gig work through gamified interface designs such as task goals, progress prompts, real-time feedback, ratings, rankings, badges, levels, and differentiated platform privileges. Although such designs are often framed as motivational devices, in platform work they are deeply entangled with algorithmic allocation, performance evaluation, income opportunities, account rights, and penalty rules. This study develops and tests a stimulus-organism-response (S-O-R) model to explain how gamified platform design shapes gig workers' continuance intention through multiple psychological mechanisms. Survey data from 338 gig workers were analyzed using partial least squares structural equation modeling (PLS-SEM). The results show that task-goal design, feedback-evaluation design, and status-honor design increase achievement, which in turn strengthens continuance intention. At the same time, feedback-evaluation and status-honor designs heighten competitive pressure, which increases emotional exhaustion and reduces continuance intention. Status-honor design also triggers loss aversion and platform dependence, thereby increasing continuance intention through a dependence-based pathway. Task-goal design increases perceived control, which further intensifies emotional exhaustion. These findings demonstrate that gamified design in platform work operates simultaneously as an engagement-enhancing mechanism and a governance mechanism. The study contributes a dual-effect perspective on gamification by showing that the same design repertoire can motivate workers while also producing soft control through comparison, dependence, and emotional depletion.

Keywords. gig economy; gamified design; platform labor; S-O-R framework; continuance intention; soft control; PLS-SEM

1. Introduction

Food delivery, ride-hailing, local logistics, online freelancing, live commerce, and other forms of platform-mediated gig work have become central to the contemporary digital economy. Digital platforms do more than match supply and demand. They structure the labor process through algorithmic dispatch, service ratings, performance

[1] Corresponding Author: Yan Jin, Business School of Dongguan City University, China. Email: jinyan@dgcu.edu.cn

indicators, account rules, and reward systems. As a result, platform workers often experience a distinctive mixture of autonomy and discipline: they may decide when to log in and which tasks to accept, yet their opportunities, earnings, and visibility are shaped by data-driven systems that continuously evaluate and rank their behavior.

Gamified design has become an important but still under-theorized component of this platform labor regime. In this study, gamified platform design refers to the embedding of points, levels, badges, rankings, task challenges, progress cues, immediate feedback, and reward-penalty rules into the management of platform work. These design elements quantify work progress, render performance comparable, and provide dynamic feedback on task completion and service quality. In ordinary consumer applications, gamification is usually discussed as a way to improve engagement. In gig platforms, however, the same elements are tied to task access, income expectations, account privileges, penalties, and reputational capital. Their effects therefore cannot be understood as merely playful or motivational.

This paper argues that gamified design in gig platforms has dual effects. On the one hand, task goals, feedback, and status recognition may enhance workers' sense of achievement and make fragmented platform work feel more purposeful. On the other hand, rankings, performance comparisons, level-based privileges, and progress demands may operate as soft control. They encourage workers to monitor themselves, protect accumulated platform assets, compete for relative standing, and adapt their work rhythm to platform-defined goals. In this sense, gamified design is both a motivational interface and a governance infrastructure.

To examine this duality, the paper develops a multi-path model based on the stimulus-organism-response (S-O-R) framework. Task-goal design, feedback-evaluation design, and status-honor design are conceptualized as platform stimuli. Achievement, competitive pressure, loss aversion, platform dependence, perceived control, and emotional exhaustion are modeled as organismic states. Continuance intention is treated as the behavioral response. Using survey data from 338 gig workers and PLS-SEM, the study tests four mechanisms: a motivational mechanism, a competition-exhaustion mechanism, a dependence formation mechanism, and a control-exhaustion mechanism.

The study makes three contributions. First, it extends research on platform labor and algorithmic management by showing how gamified interface elements translate algorithmic governance into everyday psychological experience. Second, it contributes to the gamification literature by moving beyond a one-sided engagement perspective and demonstrating that gamified design can simultaneously motivate and control. Third, it advances continuance research in platform work by distinguishing between continuance grounded in achievement and continuance grounded in dependence or loss avoidance. This distinction is important for design management because high participation metrics may conceal very different worker experiences.

2. Literature Review and Theoretical Foundation

2.1 Platform Labor and Algorithmic Management

Platform labor has evolved from a digitally mediated matching arrangement into a data-intensive mode of labor organization. Platforms rely on algorithms, ratings, time standards, and dynamic rules to coordinate dispersed workers and customers. Prior

studies show that this arrangement can produce a paradoxical form of autonomy: workers appear independent because they lack a conventional supervisor, but their economic opportunities and work rhythms are shaped by opaque algorithmic systems [1][2][3].

Algorithmic management is therefore a key lens for understanding control in platform work. Platforms allocate tasks, adjust prices, display performance information, and evaluate service quality through data systems. Workers' dependence on platform income can amplify the constraining power of these systems, especially when reputational scores, access to orders, and account privileges are linked to continuous compliance [4][5]. Existing research has emphasized dispatch algorithms, ratings, delivery deadlines, and penalty rules. Less attention has been paid to the gamified rules that accompany these systems and make control feel like progress, competition, or achievement.

2.2 Gamified Design in Platform Work

Gamification refers to the use of game elements in non-game contexts to influence motivation, experience, and behavior[6]. In organizational settings, gamified systems have been used for employee motivation, performance feedback, training, talent management, and knowledge sharing. Positive interpretations emphasize goal clarity, immediate feedback, recognition, and engagement. These features can increase workers' sense of progress and make repetitive tasks more meaningful.

Platform work differs from ordinary gamified consumer interaction. Badges, rankings, levels, challenge tasks, progress bars, deductions, and privilege rules are often integrated with labor allocation and earnings. A higher level may increase access to orders; a lower rating may reduce visibility; a missed task may threaten accumulated standing. Consequently, gamified platform design has a labor-management function. It can invite engagement while simultaneously creating performance pressure, loss sensitivity, dependence on accumulated account assets, and self-discipline.

This dual role is central to the present study. Gamified design is not treated as a neutral engagement layer added to work. It is understood as a design-management arrangement through which platforms make labor measurable, comparable, and governable. The question is not whether gamification is good or bad, but how its motivational and control effects are produced through distinct psychological paths.

2.3 Soft Control and Continuance Intention

Soft control refers to forms of organizational influence that work through norms, identity, evaluation, culture, self-monitoring, and internalized discipline rather than direct command. In platform labor, soft control is especially important because workers may not face a visible manager. Instead, they encounter goals, scores, levels, rankings, reminders, and penalties embedded in the interface. These signals can guide behavior continuously while preserving the surface appearance of autonomy.

Gamified design can therefore be understood as a carrier of soft control. Rankings and level comparisons encourage workers to watch their relative position. Progress prompts and task challenges invite workers to adjust pace and effort in response to platform-defined targets. Accumulated badges, ratings, and privileges create psychological switching costs because leaving the platform may feel like giving up

previously earned assets. Such mechanisms matter for continuance intention, which in gig work differs from conventional employee retention. Continuance may reflect enthusiasm and achievement, but it may also reflect dependence, loss avoidance, and limited alternatives.

2.4 S-O-R Framework

The stimulus-organism-response framework provides a suitable theoretical structure for explaining how platform design affects behavioral intention. Mehrabian and Russell proposed that environmental stimuli influence internal affective and cognitive states, which then shape approach or avoidance responses [7]. Jacoby further emphasized organismic states as internal appraisal processes linking external stimuli to behavior [8].

In this study, task-goal design, feedback-evaluation design, and status-honor design are the stimuli. Achievement, competitive pressure, loss aversion, platform dependence, perceived control, and emotional exhaustion represent workers' psychological responses. Continuance intention is the behavioral response. The S-O-R framework is useful because it allows gamified design to be analyzed as a set of platform stimuli that may produce both positive and negative organismic states. This is precisely the theoretical space in which the dual effect of gamified design can be specified.

3. Research Model and Hypotheses Development

3.1 Research Model

The model distinguishes three categories of gamified design. Task-goal design captures platform-defined goals, task decomposition, completion requirements, and progress cues. Feedback-evaluation design captures ratings, rankings, user evaluations, income feedback, and real-time performance prompts. Status-honor design captures levels, badges, titles, identity labels, and differentiated platform privileges. These design categories are not merely interface functions; they are ways through which platforms present standards, outcomes, and status differences to workers.

Six organismic states are included. Achievement refers to workers' positive experience of task completion, recognition, and level advancement. Competitive pressure captures tension generated by comparison, ranking, and performance evaluation. Loss aversion reflects sensitivity to losing accumulated levels, ratings, privileges, or future opportunities. Platform dependence captures attachment to platform resources and account-based opportunities. Perceived control reflects the subjective sense that one's work rhythm is continuously guided by platform rules. Emotional exhaustion reflects the depletion of psychological resources under sustained pressure.

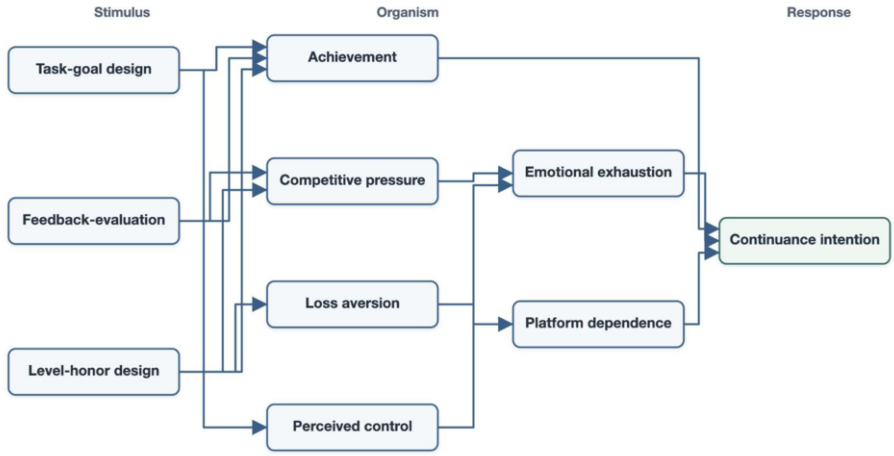


Figure 1. Conceptual Model Based on the S-O-R Framework

3.2 Motivational Mechanism

The motivational mechanism captures the engagement-enhancing side of gamified design. Task goals, progress indicators, feedback, rankings, and status recognition can make effort visible and meaningful. Because gig work is often fragmented, repetitive, and weakly embedded in a stable organizational career path, achievement can become an important psychological resource. Workers who perceive that their task completion, service quality, and level advancement are recognized by the platform are more likely to continue participating.

Accordingly, task-goal design, feedback-evaluation design, and status-honor design are expected to increase achievement, and achievement is expected to strengthen continuance intention. This pathway represents the positive design promise of gamification: it can support engagement by making progress legible and by converting dispersed labor into recognized accomplishment.

H1: Gamified platform design enhances gig workers' continuance intention through achievement.

3.3 Competition-Exhaustion Mechanism

Gamified design also introduces comparison. Feedback-evaluation and status-honor designs place workers in systems where scores, rankings, badges, and levels communicate relative standing. When these signals are linked to order access, income expectations, platform privileges, or penalty risks, they may generate competitive pressure. Workers must not only complete tasks but also keep pace with constantly visible performance benchmarks.

Competitive pressure can consume psychological resources. Workers exposed to continuous evaluation, comparison, and performance chasing may experience platform work less as flexible participation and more as an ongoing race. Emotional exhaustion

is therefore expected to mediate the negative consequences of competitive pressure for continuance intention.

H2: Feedback-evaluation design and status-honor design reduce continuance intention through competitive pressure and emotional exhaustion.

3.4 Dependence Formation Mechanism

Status-honor design may also create dependence. Levels, badges, titles, ratings, credit records, and differentiated privileges are cumulative. Workers obtain these resources through continued task completion, favorable evaluations, and compliance with platform rules. Leaving the platform may therefore be perceived as losing accumulated assets that are difficult to transfer elsewhere.

Loss aversion provides the psychological basis for this mechanism. Workers who are sensitive to losing accumulated status and privileges may become more dependent on the platform, and this dependence may increase continuance intention. Importantly, this path does not necessarily indicate positive identification with the platform. It may instead reflect a defensive intention to preserve existing resources and avoid loss.

H3: Status-honor design increases continuance intention through loss aversion and platform dependence.

3.5 Control-Exhaustion Mechanism

Task-goal design may improve clarity while also increasing perceived control. Platform-defined task goals, progress reminders, limited-time challenges, and completion requirements help workers understand what the platform expects. Yet these goals are usually generated by platform rules rather than by workers themselves. To obtain rewards or avoid penalties, workers may need to adjust task acceptance, pace, and service behavior in response to platform prompts.

Perceived control is expected to intensify emotional exhaustion because workers must continually attend to goals, cues, and performance requirements defined by the platform. This mechanism captures the soft-control side of goal design: the interface does not simply inform workers; it can also direct their attention, rhythm, and emotional labor.

H4: Task-goal design increases emotional exhaustion through perceived control.

All hypotheses of this study and their corresponding structural paths are summarized in Table 1.

Table 1. Research hypotheses and corresponding structural paths

Hypothesis	Mechanism	Structural path	Theoretical meaning
H1a	Motivational mechanism	Task-goal design -> Achievement	Task goals, progress cues, and task decomposition enhance workers' sense of task accomplishment.
H1b	Motivational mechanism	Feedback-evaluation design -> Achievement	Ratings, evaluations, income cues, and real-time feedback allow workers to perceive that their effort is recorded and recognized.
H1c	Motivational mechanism	Status-honor design -> Achievement	Levels, badges, titles, and differentiated rights provide symbolic recognition and identity distinction.

H1d	Motivational mechanism	Achievement -> Continuance intention	Achievement, as a positive psychological state, strengthens workers' intention to continue platform work.
H2a	Competition-exhaustion mechanism	Feedback-evaluation design -> Competitive pressure	Ratings, rankings, and user evaluations place workers in a continuous state of being evaluated and compared.
H2b	Competition-exhaustion mechanism	Status-honor design -> Competitive pressure	Levels, badges, and titles create status distinctions and relative comparison among workers.
H2c	Competition-exhaustion mechanism	Competitive pressure -> Emotional exhaustion	Sustained competition and performance chasing consume workers' psychological resources.
H2d	Competition-exhaustion mechanism	Emotional exhaustion -> Continuance intention	Emotional exhaustion, as a negative psychological state, weakens workers' intention to continue platform work.
H3a	Dependence formation mechanism	Status-honor design -> Loss aversion	Accumulated levels, evaluations, and platform rights increase workers' sensitivity to potential loss.
H3b	Dependence formation mechanism	Loss aversion -> Platform dependence	Concerns about losing accumulated platform resources increase workers' dependence on the platform.
H3c	Dependence formation mechanism	Platform dependence -> Continuance intention	Platform dependence raises exit costs and makes workers more likely to remain active.
H4a	Control-exhaustion mechanism	Task-goal design -> Perceived control	Platform-defined task goals, progress prompts, and completion requirements may make workers feel continuously guided.
H4b	Control-exhaustion mechanism	Perceived control -> Emotional exhaustion	Perceived control increases psychological burden and further intensifies emotional exhaustion.

4. Methodology

4.1 Sample and Data Collection

Data were collected through a questionnaire survey of gig workers engaged in platform-mediated work, including food delivery, ride-hailing, logistics and local services, online freelancing, and other digital labor contexts. The questionnaire focused on workers' perceptions of gamified platform design, psychological responses to platform rules, and continuance intention.

A total of 386 questionnaires were collected. After removing incomplete responses, invalid questionnaires, and responses with obvious bias, 338 valid questionnaires remained, yielding an effective response rate of 87.6%. The final sample covered different genders, ages, education levels, and platform types, providing a

suitable empirical basis for examining variation in workers' perceptions of gamified design.

4.2 Measurement Development

All constructs were measured using five-point Likert scales. Task-goal design, feedback-evaluation design, and status-honor design were modeled as exogenous stimulus variables. Achievement, competitive pressure, loss aversion, platform dependence, perceived control, and emotional exhaustion were modeled as organismic variables. Continuance intention was modeled as the response variable. Each latent construct was measured reflectively using five items.The latent variables and their corresponding measurement item codes are summarized in Table 2.

Table 2. Constructs and measurement indicators

Construct	Number of items	Code
Task-goal design	5	A1-A5
Feedback-evaluation design	5	B1-B5
Status-honor design	5	C1-C5
Achievement	5	E1-E5
Competitive pressure	5	F1-F5
Loss aversion	5	G1-G5
Platform dependence	5	H1-H5
Perceived control	5	I1-I5
Emotional exhaustion	5	J1-J5
Continuance intention	5	K1-K5

4.3 Analytical Procedure

The empirical analysis used partial least squares structural equation modeling (PLS-SEM). PLS-SEM is appropriate for this study because the model includes multiple mediated paths and focuses on explaining and predicting continuance intention. The analysis proceeded in three stages. First, descriptive statistics were examined. Second, the measurement model was assessed through reliability, convergent validity, discriminant validity, and collinearity diagnostics, including Cronbach's alpha, rho_A, composite reliability, average variance extracted (AVE), HTMT, and VIF checks. Third, the structural model was evaluated using path coefficients, t values, p values, R2, mediation tests, PLSpredict, Q2 predict, and CVPAT.

The analytical logic follows the dual-effect premise of the paper. Rather than estimating a single net effect of gamified design, the model tests whether different design components activate achievement, competition, dependence, and control pathways simultaneously.

5. Results

5.1 Descriptive Statistics

Construct means ranged from 3.786 to 3.927, and standard deviations ranged from 0.814 to 0.894. All construct means were above the midpoint of the five-point scale, indicating that respondents' perceptions of platform gamified design and related psychological states were moderately high. Continuance intention was strongly associated with platform dependence and achievement, while emotional exhaustion was negatively associated with continuance intention.The descriptive statistics for each variable construct are presented in Table 3.

Table 3. Descriptive statistics

Construct	Mean	Standard deviation
Task-goal design	3.821	0.882
Feedback-evaluation design	3.864	0.893
Status-honor design	3.852	0.884
Achievement	3.891	0.831
Competitive pressure	3.804	0.893
Loss aversion	3.927	0.814
Platform dependence	3.829	0.831
Perceived control	3.822	0.894
Emotional exhaustion	3.786	0.880
Continuance intention	3.818	0.834

5.2 Measurement Model Assessment

The measurement model showed acceptable standardized outer loadings. Cronbach's alpha, rho_A, and composite reliability all exceeded conventional thresholds, and all AVE values were above 0.50, indicating internal consistency and convergent validity. HTMT values did not exceed the 0.85 and 0.90 thresholds, supporting discriminant validity. Outer and inner VIF values did not indicate serious collinearity problems. Overall, the measures were reliable and valid for structural model testing.The results of composite reliability and convergent validity tests for each variable are shown in Table 4.

Table 4. Reliability and convergent validity

Construct	α	rho_A	rho_C	AVE
Task-goal design	0.881	0.888	0.913	0.677
Feedback-evaluation design	0.894	0.896	0.922	0.702

Status-honor design	0.887	0.890	0.917	0.689
Achievement	0.869	0.871	0.905	0.656
Competitive pressure	0.877	0.882	0.911	0.671
Loss aversion	0.865	0.869	0.903	0.650
Platform dependence	0.857	0.861	0.897	0.635
Perceived control	0.892	0.895	0.920	0.698
Emotional exhaustion	0.883	0.887	0.915	0.682
Continuance intention	0.860	0.861	0.900	0.642

5.3 Structural Model Assessment

The structural model achieved acceptable fit for a PLS-SEM analysis. SRMR was 0.066, below the 0.08 criterion, and NFI was 0.842, indicating an acceptable overall model fit. The R2 values showed that the model explained meaningful variance in continuance intention, emotional exhaustion, achievement, platform dependence, and the other endogenous constructs. All hypothesized direct paths were statistically significant.

The motivational mechanism was supported: task-goal design, feedback-evaluation design, and status-honor design significantly increased achievement, and achievement significantly increased continuance intention. The competition-exhaustion mechanism was also supported: feedback-evaluation design and status-honor design increased competitive pressure, competitive pressure increased emotional exhaustion, and emotional exhaustion reduced continuance intention. The dependence formation mechanism was supported: status-honor design increased loss aversion, loss aversion increased platform dependence, and platform dependence increased continuance intention. Finally, the control-exhaustion mechanism was supported: task-goal design increased perceived control, and perceived control increased emotional exhaustion.The structural path coefficients, t-values, p-values, and test conclusions for all hypotheses are presented in Table 5.

Table 5. Structural path test results

Hypothesis	Path	β	t	p	Conclusion
H1a	Task-goal design -> Achievement	0.152	2.201	0.028	Supported
H1b	Feedback-evaluation design -> Achievement	0.244	3.872	<0.001	Supported
H1c	Status-honor design -> Achievement	0.240	3.720	<0.001	Supported
H1d	Achievement -> Continuance intention	0.407	9.807	<0.001	Supported
H2a	Feedback-evaluation design -> Competitive pressure	0.204	3.303	0.001	Supported
H2b	Status-honor design -> Competitive pressure	0.313	5.083	<0.001	Supported
H2c	Competitive pressure -> Emotional exhaustion	0.381	8.537	<0.001	Supported
H2d	Emotional exhaustion -> Continuance intention	-0.200	4.714	<0.001	Supported

H3a	Status-honor design -> Loss aversion	0.467	10.619	<0.001	Supported
H3b	Loss aversion -> Platform dependence	0.504	12.315	<0.001	Supported
H3c	Platform dependence -> Continuance intention	0.417	10.743	<0.001	Supported
H4a	Task-goal design -> Perceived control	0.379	8.003	<0.001	Supported
H4b	Perceived control -> Emotional exhaustion	0.315	5.920	<0.001	Supported

5.4 Mediation Analysis

The mediation results further confirmed the multi-path logic of the model. Task-goal design, feedback-evaluation design, and status-honor design indirectly increased continuance intention through achievement. Feedback-evaluation design and status-honor design indirectly reduced continuance intention through competitive pressure and emotional exhaustion. Task-goal design indirectly reduced continuance intention through perceived control and emotional exhaustion. Status-honor design also indirectly increased continuance intention through loss aversion and platform dependence. These results show that continuance can be produced by both positive engagement and dependence-based soft control.The calculated results of the specific indirect effects for each path are shown in Table 6.

Table 6. Specific indirect effects

Indirect path	Effect	p
Task-goal design -> Achievement -> Continuance intention	0.062	0.034
Feedback-evaluation design -> Achievement -> Continuance intention	0.099	<0.001
Status-honor design -> Achievement -> Continuance intention	0.097	0.001
Feedback-evaluation design -> Competitive pressure -> Emotional exhaustion -> Continuance intention	-0.016	0.014
Status-honor design -> Competitive pressure -> Emotional exhaustion -> Continuance intention	-0.024	0.002
Task-goal design -> Perceived control -> Emotional exhaustion -> Continuance intention	-0.024	0.001
Status-honor design -> Loss aversion -> Platform dependence -> Continuance intention	0.098	<0.001

5.5 Predictive Assessment

PLSpredict results indicated that all endogenous constructs had positive Q2 predict values, demonstrating predictive relevance. CVPAT further showed that the model's overall predictive performance was superior to both the indicator-mean benchmark and the linear model benchmark. However, the advantage over the linear benchmark was not significant for some constructs; therefore, the predictive interpretation should remain appropriately cautious.The goodness-of-fit R^2 values for the endogenous variables and the PLSpredict predictive indicators are presented in Table 7.

Table 7. R2 and PLSpredict results

Endogenous construct	R^2	Adjusted R^2	Q^2 predict
Achievement	0.302	0.295	0.286
Competitive pressure	0.219	0.214	0.208
Perceived control	0.144	0.141	0.137
Emotional exhaustion	0.321	0.317	0.115
Loss aversion	0.218	0.216	0.212
Platform dependence	0.254	0.251	0.073
Continuance intention	0.402	0.396	0.105

6. Discussion

The findings support the central argument that gamified platform design is a dual-effect design-management mechanism. It does not merely add enjoyable or motivating features to platform work. It also organizes workers' attention, comparison, dependence, and self-discipline. This duality is visible in the coexistence of positive and negative paths. Achievement strongly increased continuance intention (beta = 0.407), suggesting that gamified design can make platform work more meaningful by converting task completion and performance feedback into recognizable progress. At the same time, emotional exhaustion reduced continuance intention (beta = -0.200), showing that the same design environment can undermine continued participation when it intensifies competition and perceived control.

The first theoretical implication concerns gamification research. Much of the managerial discourse around gamification emphasizes engagement, motivation, and performance. This study does not reject that view; rather, it qualifies it. Task-goal, feedback-evaluation, and status-honor designs significantly enhanced achievement, and achievement was a powerful predictor of continuance. However, the results also show that gamified design becomes a governance mechanism when it is linked to rankings, ratings, levels, and account privileges. In platform labor, gamification is not simply a design layer for experience; it is part of the infrastructure through which work is measured and governed.

The second implication concerns algorithmic management and platform labor. Prior research has shown how platforms govern workers through dispatch algorithms, ratings, information asymmetries, and dynamic rules. This study adds that gamified design translates algorithmic management into everyday psychological cues. Workers may not experience control only as a formal rule or penalty. They may experience it as a task target, a progress bar, a leaderboard, a level threshold, or a potential loss of accumulated privilege. Soft control is therefore not external to design; it is enacted through design.

The third implication concerns continuance intention. The positive effect of platform dependence on continuance intention (beta = 0.417) was slightly stronger than the effect of achievement (beta = 0.407). This finding is theoretically important because

it suggests that continued participation should not be interpreted automatically as satisfaction or commitment. Some workers may continue because the platform provides achievement and recognition, whereas others may continue because accumulated ratings, levels, and privileges increase switching costs. For design management and information systems research, this distinction warns against treating retention metrics as unambiguous evidence of user value.

The fourth implication concerns the S-O-R framework. The results demonstrate that a single class of platform stimuli can activate multiple organismic states with different behavioral consequences. Achievement represents a positive organismic state; competitive pressure, perceived control, and emotional exhaustion represent strain-related states; loss aversion and platform dependence represent dependence-related states. This multi-path specification provides a more nuanced explanation of how digital interface design shapes labor behavior than a simple positive or negative model of gamification.

From a practical perspective, platform firms should design gamification with transparency, proportionality, and worker well-being in mind. Goal prompts and feedback can support clarity, but platforms should avoid excessive reliance on ranking competition and punitive comparison. Level systems and badges should not obscure the real consequences of account status for task access and income opportunities. Workers should be able to understand how points, scores, rankings, deductions, and privileges are calculated, and they should have meaningful channels for appeal and correction. Regulators and labor advocates should also attend to interface-level governance, because soft control may be embedded in design features that appear benign or playful.

7. Conclusion and Implications

This study examined how gamified design affects gig workers' continuance intention through multiple psychological paths. Based on survey data from 338 gig workers and PLS-SEM analysis, the study found that gamified design simultaneously promotes achievement, generates competitive pressure, fosters loss-based platform dependence, and increases perceived control. These mechanisms jointly shape continuance intention. The core conclusion is that gamified platform design is both motivation and soft control.

For theory, the study contributes a dual-effect perspective that integrates gamification, platform labor, and S-O-R research. It shows that gamified design can enhance engagement while also producing governance effects through comparison, dependence, and emotional depletion. For design management, the study suggests that responsible platform design should evaluate not only whether gamification increases activity, but also how it redistributes psychological burden and exit costs. For information systems research, the findings demonstrate the value of examining interface design as a sociotechnical governance mechanism rather than as a purely experiential feature.

For practice, platforms should preserve the constructive side of gamification--clear goals, timely feedback, and meaningful recognition--while reducing mechanisms that create excessive pressure or hidden dependence. Transparent rules, less punitive ranking systems, portable reputation mechanisms, and effective appeals processes can help ensure that gamified design supports sustainable participation rather than simply extracting more labor through soft control.

References

[1] Rosenblat, A., & Stark, L. (2016). Algorithmic labor and information asymmetries: A case study of Uber's drivers. International Journal of Communication, 10, 3758-3784.

[2] Vallas, S., & Schor, J. B. (2020). What do platforms do? Understanding the gig economy. Annual Review of Sociology, 46, 273-294.

[3] Wood, A. J., Graham, M., Lehdonvirta, V., & Hjorth, I. (2019). Good gig, bad gig: Autonomy and algorithmic control in the global gig economy. Work, Employment and Society, 33(1), 56-75.

[4] Schor, J. B., Attwood-Charles, W., Cansoy, M., Ladegaard, I., & Wengronowitz, R. (2020). Dependence and precarity in the platform economy. Theory and Society, 49, 833-861.

[5] Sun, P. (2019). Your order, their labor: An exploration of algorithms and laboring on food delivery platforms in China. Chinese Journal of Communication, 12(3), 308-323.

[6] Seaborn, K., & Fels, D. I. (2015). Gamification in theory and action: A survey. International Journal of Human-Computer Studies, 74, 14-31.

[7] Mehrabian, A., & Russell, J. A. (1974). An approach to environmental psychology. MIT Press.

[8] Jacoby, J. (2002). Stimulus-organism-response reconsidered: An evolutionary step in modeling consumer behavior. Journal of Consumer Psychology, 12(1), 51-57.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261132

Modified Ratios for Assessing the Impact of Artificial Intelligence on the Economic Efficiency of Industrial Enterprises

Plamena ZLATEVA[a,b,1], Marin GALABOV[a] and Dimiter VELEV[a]
[a]*University of National and World Economy, Sofia, Bulgaria*
[b]*Institute of Robotics, Bulgarian Academy of Sciences, Sofia, Bulgaria*
ORCiD ID: Dimiter Velev https://orcid.org/0000-0003-3030-1819
Plamena Zlateva https://orcid.org/0000-0002-0153-5811

Abstract. Artificial intelligence (AI) is increasingly being established as an important factor in increasing the productivity, efficiency and competitiveness of modern industrial enterprises. Classic efficiency ratios cannot directly assess the different aspects of the impact of AI implementation on the overall activity of an enterprise. This circumstance creates a methodological gap between the processes of digital transformation with AI and the existing tools for assessing the economic efficiency of enterprises. The aim of the paper is to propose modified efficiency ratios for assessing the impact of artificial intelligence on the economic efficiency of industrial enterprises. The developed conceptual methodological framework extends four classic ratios for economic efficiency in term of total costs, production cost, personnel and capital, by including specific variables taking into account the impact of AI on the productivity of an industrial enterprise.

Keywords. Artificial Intelligence, Economic Efficiency, Industrial Enterprises, Modified economic efficiency ratios

1. Introduction

Nowadays, the accelerated development of digital technologies and artificial intelligence (AI) is transforming industrial enterprises worldwide [1, 2]. Artificial intelligence, advanced automation, and intelligent decision-making are among the most significant technologies that determine the sustainable development and global competitiveness of every enterprise [3, 4, 5]. The development of Industry 4.0 and the transition to Industry 5.0 significantly increase the role of AI in production, logistics, marketing, strategic management, etc. [6, 7, 8]. Modern machine learning algorithms, computer vision systems, and intelligent robotics are increasingly integrated into the production and management systems of industrial enterprises, creating opportunities for significantly increasing their efficiency [9, 10, 11, 12].

The recent analyses of the activities of the efficient industrial enterprises show that the AI implementation contributes to higher productivity, better resource utilization,

[1] Corresponding Author: Plamena Zlateva, University of National and World Economy, 19 December 8th Str., 1700 Sofia, Bulgaria; E-mail: plamzlateva@unwe.bg.

higher quality of production, as well as leading to better organizational adaptability, higher added value, and greater profits [13, 14, 15].

It is important to note that widespread implementation of AI in the activities of industrial enterprises requires the development of appropriate analytical measures to assess the impact of AI on their the economic efficiency and sustainable development [16, 17, 18]. It is well known that classic economic efficiency ratios, established by theory and practice, are a useful tool for analyzing and assessing the activities and sustainability of industrial enterprises [19, 20, 21]. However, it is important to note that these ratios do not explicitly assess the impact of artificial intelligence on the economic efficiency of industrial enterprises such as cost savings, increased revenue, energy optimization, improved productivity, higher quality, intelligent automation and more. Therefore, modified indicators are needed that more accurately assess the economic effects of implementing AI in industrial enterprises.

Recent research has increasingly focused on assessing the economic value created by the implementation of artificial intelligence in enterprises. Existing approaches primarily evaluate this value through return-on-investment (ROI) analysis, productivity indicators, business value creation models and digital transformation metrics. Recent academic studies, together with reports published by international organizations such as the Organisation for Economic Co-operation and Development (OECD), demonstrate that the implementation of artificial intelligence contributes to productivity growth, innovation, resource optimization and improvements in enterprise competitiveness [22–28]. However, these approaches assess the overall economic contribution of artificial intelligence without explicitly incorporating its effects into the traditional system of economic efficiency ratios widely used in enterprise analysis. Consequently, there remains a methodological gap regarding the adaptation of classical efficiency indicators to the conditions of artificial intelligence implementation in industrial enterprises.

In this context, the present study investigates how the classical economic efficiency ratios can be extended to explicitly account for the economic effects of artificial intelligence implementation. The proposed conceptual methodological framework preserves the economic interpretation of the traditional indicators while incorporating additional variables reflecting the impact of artificial intelligence on enterprise revenues, costs and resource utilization.

The aim of the paper is to propose four modified efficiency ratios for assessing the impact of AI on the economic efficiency of industrial enterprises. The proposed modified efficiency ratios include additional variables that are directly related to AI implementation. Therefore, these coefficients can be considered as a useful methodological framework for assessing the impact of artificial intelligence on the economic efficiency of industrial enterprises and thus help managers make more informed and adequate decisions for the sustainable development of their business.

It is necessary to point out that this study adopts a conceptual analytical approach. First, four widely used classical economic efficiency ratios were selected because they represent the main categories of enterprise resources, namely total costs, production costs, personnel costs and capital. Second, the variables reflecting the effects of artificial intelligence were identified on the basis of the economic mechanisms underlying the impact of artificial intelligence on enterprise resource utilization and economic efficiency. Finally, these variables were incorporated into the classical efficiency ratios according to their economic interpretation. Variables representing additional revenue resulting from the implementation of artificial intelligence are added to the numerator,

whereas variables reflecting cost savings are deducted from the denominator. Conversely, variables representing expenditures and investments associated with the implementation of artificial intelligence are added to the denominator. The proposed methodological framework is conceptual in nature and requires future empirical validation using accounting, production and management data from industrial enterprises.

2. Analysis of the economic efficiency of an industrial enterprise through classic ratios

This study examines four classic ratios for analyzing and evaluating the economic efficiency of an industrial enterprise.

2.1. Efficiency ratio of total costs

The efficiency ratio of total costs measures the amount of net revenue generated from sales of production per total costs over a certain period:

$$ER_{TC} = Efficiency\ ratio\ of\ total\ costs = \frac{Net\ sales\ revenue}{Total\ costs} \quad (1)$$

Here, a high value of the ratio indicates high efficiency of the industrial enterprise, and a low value indicates excessive costs or low productivity industry enterprise.

2.2. Efficiency ratio of production costs

The efficiency ratio of production cost measures the amount of net revenue from sales of production generated per unit of cost of sold production during a given period:

$$ER_{PC} = Efficiency\ ratio\ of\ production\ costs = \frac{Net\ sales\ revenue}{Cost\ of\ sold\ production} \quad (2)$$

This ratio indicates the efficiency of the production process, the degree of optimization of production costs, the quality of production organization. It is used to evaluate rational utilization of raw materials, labor productivity, technological efficiency, losses caused by defects and downtime.

2.3. Personnel Efficiency Ratio

The personnel efficiency ratio measures the relationship between net revenue generated from production sales and personnel productivity over a certain period:

$$ER_{AP} = Personnel\ efficiency\ ratio = \frac{Net\ sales\ revenue}{Personnel\ costs} \quad (3)$$

This ratio is used to assess labor productivity, the level of personnel qualification, the effectiveness of the production organization and personnel management. A higher value of the ratio usually indicates better personnel management, higher labor productivity and greater overall operational efficiency.

2.4. Capital efficiency ratio

The capital efficiency ratio measures the ability of an industrial enterprise's total capital to generate revenue from sales of production over a certain period:

$$ER_C = Capital\ efficiency\ ratio = \frac{Sales\ revenue}{Total\ capital} \quad (4)$$

The ratio shows the usefulness of total capital investments, the degree of utilization of the financial resources of the industrial enterprise, the production return on total capital. This ratio is important for investors, financial analysts, creditors, strategic management.

3. Modified ratios for assessing the economic efficiency of industrial enterprises with implemented artificial intelligence

The idea is to propose modified coefficients for assessing the economic efficiency of industrial enterprises with implemented AI.

For this reason, it is necessary to define the specific variables that directly take into account the impact of the application of AI for each of the basic ratios of the economic efficiency (1)-(4).

All AI-related adjustment variables, are measured in monetary units (EUR). Consequently, the modified efficiency ratios express the amount of sales revenue generated per adjusted monetary unit of the corresponding resource after accounting for the economic effects of AI implementation

Here, the modified ratios for assessing the impact of artificial intelligence on the economic efficiency of an industrial enterprise are defined as follows.

3.1. Modified efficiency ratio of total costs for AI impact assessment

The modified efficiency ratio of total costs for AI impact assessment is defined as:

$$ER_{TC}^{AI} = \frac{Net\ sales\ revenue + AI_R}{Total\ costs + AI_C - AI_S}, \quad (5)$$

where: AI_R is the additional revenue generated through AI;

AI_C - the operating costs for implementing and maintaining AI;

AI_S - the total cost savings due to AI implementation.

This modified efficiency ratio shows how the implementation of AI leads to a real increase in the efficiency of the enterprise's total costs. More specifically, it considers whether the additional revenue and reduced total costs from using AI compensate the additional costs of implementing and maintaining the AI itself. This ratio can be used to measure an enterprise's ability to generate revenue and cost savings through the AI use.

The variable AI_R is associated with increased productivity of the industrial enterprise, higher quality of production, better customer service and more accurate forecasting of market demand. The implemented AI contributes to better production planning, reducing order fulfillment times and increasing production capacity without a significant increase in the resources used. The value of the AI_R variable is associated with an increase in sales

volume, an increase in market share, a decrease in defective products, the creation of new products through AI, and an improvement in customer satisfaction. The higher the AI_R value, the greater the contribution of AI to the enterprise's revenue.

The variable AI_R appears in the numerator of all modified ratios because additional revenue generated through AI implementation affects the overall economic performance of the enterprise regardless of whether efficiency is assessed relative to costs, production resources, personnel resources or capital resources. The variable AI_C represents operating costs associated with implementing and maintaining AI. Although AI leads to a reduction in some costs and generation of additional revenue, its functioning requires operational investments in software systems, maintenance of special hardware, licenses, model updates, data storage, specialized personnel, and more. The value of the variable AI_C depends on the value of license fees, cloud services costs, data processing and storage costs, AI specialists' salaries, model update and training costs, and more.

The variable AI_S shows the cost savings due to AI implementation. These savings can be the result of automation of routine activities, optimization of production processes, better use of resources and reduction of administrative costs. In the practice of industrial enterprises, AI often leads to higher productivity with the same amount of resources used. The value of the variable AI_S variable depends on the reduction of labor costs, reduction of logistics costs, optimization of inventory, reduction of operating costs, and reduction of time for performing production operations. In essence, the variable AI_S measures the direct financial impact of intelligent automation on the cost structure of an industrial enterprise.

It is necessary to point out that the modified efficiency ratio (5) is valid only when the adjusted denominator remains positive. Therefore, the value of AIS should not exceed the corresponding baseline total cost category from which it is deducted.

3.2. Modified efficiency ratio of production costs for AI impact assessment

The modified efficiency ratio of production costs for AI impact assessment is defined as:

$$ER_{PC}^{AI} = \frac{Net\ sales\ revenue + AI_R}{Cost\ of\ sold\ production + AI_P - AI_S - AI_E - AI_Q}, \quad (6)$$

where: AI_R is the additional revenue generated through AI;

AI_S - the total cost savings due to AI implementation.

AI_E - the energy cost savings through AI;

AI_Q - the reduction of quality-related losses through AI;

AI_P – the production-oriented AI operating costs.

Modified efficiency ratio of production costs assesses the impact of artificial intelligence on production efficiency through changes in the cost of production. In modern industrial enterprises, AI is used for optimization of production parameters, intelligent control of machines, automated monitoring of technological processes and predictive maintenance of equipment.

Through real-time analysis of production data, AI can minimize raw material costs, reduce energy consumption, and prevent defects from occurring early in the production process. This is why savings in materials, energy, and scrap costs are deducted from the cost price in the denominator of the coefficient. However, the use of specialized production AI systems generates additional ongoing costs that must be accounted for through the variable AI_P.

The variable AI_E represents the energy cost savings achieved through the implementation of AI in production management systems. Through continuous analysis of equipment operating parameters, intelligent algorithms can optimize machine loads, minimize energy losses, and reduce peak electricity consumption. The value of the variable AI_E depends on the degree of automated energy consumption management, the number of energy-intensive processes, the accuracy of forecasting algorithms, and the possibilities for optimizing production load. This variable is especially important in energy-intensive industries such as metallurgy, chemical industry, and mechanical engineering.

The variable AI_Q, associated with reduction of quality losses through AI, essentially represents a reduction in costs caused by production waste, defective products, complaints, technological deviations, and others. For example, machine vision, big data analytics, intelligent quality control systems, and early defect detection algorithms enable timely identification and resolution of problems in the production process. The value of the variable AI_Q depends on the percentage of reduced production waste, reduced rework costs, reduced warranty service costs, reduced customer complaints, and improved stability of the production process. This variable shows the contribution of AI to improving quality and reducing production losses.

The variable AI_P represents only the operating costs for implementing, operating and maintaining artificial intelligence systems, which are directly related to the production process of the industrial enterprise and participate in the formation of the cost of production, unlike the variable AI_C (the operating costs for implementing and maintaining AI), which may include administrative, marketing or management applications of AI. The value of the variable AI_P may include costs for licensing of production AI platforms, use of machine vision systems and automated quality control, predictive maintenance of equipment, processing and analysis of big production data, training and adapting AI models to specific production, technical support of intelligent production systems and others.

It is necessary to point out that the modified efficiency ratio (6) is valid only when the adjusted production cost denominator remains strictly positive:

$$Cost\ of\ sold\ production + AI_P\ - AI_S - AI_{\mathrm{E}} - AI_Q > 0.$$

This condition ensures the economic interpretability and mathematical validity of the modified efficiency ratio.

3.3. Modified personnel efficiency ratio for AI impact assessment

The modified personnel efficiency ratio for AI impact assessment is defined as:

$$ER_{AP}^{AI} = \frac{Net\ sales\ revenue + AI_R}{Personnel\ costs - AI_L + AI_T}, \tag{7}$$

where: AI_L is the personnel cost savings due to AI implementation;

AI_T - the costs of training personnel to work with AI systems

The variable AI_L reflects the personnel cost savings generated through the implementation of artificial intelligence. Such savings may result from the automation of repetitive tasks, improved workforce planning, enhanced labor productivity, and reduced reliance on manual operations. Higher values of AI_L indicate a stronger

economic effect of AI in reducing labor-related expenditures while maintaining or increasing production output.

The variable AI_T takes into account the costs of training personnel to work with AI systems. This is necessary because the successful implementation of AI requires the acquisition of new digital competencies, retraining and upgrading of personnel.

This modified ratio is designed to measure the impact of intelligent automation with AI on labor productivity. Unlike the classic ratio, which considers only human labor as a factor for creating value, the modified version of the ratio also takes into account the impact of applied AI in the performance of production and management functions. The modified ratio takes into account the increase in net sales revenue and the decrease in the average number of personnel in the industrial enterprise, due to the improvement of production with automated systems with AI. In addition, the costs of training and retraining of personnel to acquire skills for working with intelligent systems are also included. Therefore, the ratio measures both the quantitative reduction and the qualitative development of personnel in the conditions of digital transformation.

3.4. Modified Capital efficiency ratio for AI impact assessment

The modified capital efficiency ratio for AI impact assessment is defined as:

$$ER_C^{AI} = \frac{Sales\ revenue + AI_R}{Total\ capital + AI_I - AI_S} \quad (8)$$

where: AI_I are the enterprise's capital investments in artificial intelligence assets.

The variable AI_I shows that in modern industrial enterprises, investments in AI represent a specific form of capital investment aimed at increasing the efficiency of production and management systems. In particular, investments in intelligent production systems, industrial robots, machine vision, digital twins and analytical platforms increase the value of the capital employed. Therefore, these investments are added to the denominator of the coefficient.

On the other hand, AI increases the efficiency of production capacity utilization by reducing downtime, improving equipment utilization and extending its service life. This results in additional revenues AI_R, which are included in the numerator.

At the same time, it is necessary to take into account the cost savings due to AI implementation AI_S. The subtraction of variable AI_S from the denominator reflects the reduction in capital requirements as a result of the optimization of the enterprise resources through AI implementation. In economic terms, the savings generated by AI improve the utilization efficiency of the employed capital and therefore reduce the effective capital base required to generate a given level of revenue.

The higher value of the modified efficiency ratio indicates that investments in AI generate a higher return on total capital and improve the overall financial efficiency of the industrial enterprise.

4. Discussion

The modified ratios developed in this study represent an attempt to adapt classic methods for analyzing economic efficiency to the conditions of digital transformation and the implementation of AI in industrial enterprises. The main advantage of the proposed

conceptual methodological framework is that it preserves the economic logic of the classical efficiency ratios, while incorporating additional variables reflecting the economic effects of AI implementation . The proposed modified ratios enable more precise accounting of additional revenues, realized cost savings, reduced energy consumption, improved product quality, and investments in AI. This creates an opportunity for a more realistic assessment of the impact of AI on the economic efficiency of industrial enterprises compared to classic ratios that do not explicitly consider such factors.

The proposed methodological framework complements the existing approaches for assessing the economic impact of artificial intelligence described in the recent scientific literature. Most contemporary studies evaluate the economic contribution of artificial intelligence through productivity indicators, return-on-investment measures, business value creation models, digital transformation metrics and performance measurement systems. Although these approaches provide valuable information regarding the overall economic effects of artificial intelligence, they are not directly integrated into the traditional system of the economic efficiency ratios widely applied in enterprise financial analysis. The modified ratios proposed in this study preserve the economic interpretation of the classical efficiency indicators while extending their analytical capabilities by explicitly incorporating the main economic effects of artificial intelligence implementation.

It is necessary to point out that determining the values of some of the defined variables is associated with certain methodological difficulties. Measuring the additional revenues generated as a result of the implementation of AI, as well as determining the savings achieved through intelligent automation, requires the availability of an appropriate information base and reliable methods for quantitative assessment.

In many cases, the economic effects of implementing AI in industrial enterprises are manifested indirectly through increasing production quality, improving customer service, reducing production downtime, optimizing resource utilization or reducing operational risk, which makes their direct measurement difficult. Another important aspect is the sectoral specificity of industrial enterprises. The impact of AI on the economic efficiency may differ significantly between machine-building, automotive, chemical, metallurgical or other manufacturing sectors. Therefore, the practical application of the defined additional variables should take into account the characteristics of the specific production process and the degree of implementation of smart technologies.

Nowadays, a realistic approach for assessing the impact of AI on the economic efficiency is to estimate the values of the defined AI variables using accounting, production and management information systems already available in industrial enterprises.. In most practical applications, the values of the modified ratios may be determined by comparing enterprise indicators before and after the implementation of AI or by comparing similar production processes operating with and without AI support. Depending on the available information, forecasting models and benchmarking techniques may also be applied to estimate the expected values of the indicators in the absence of artificial intelligence.

The practical estimation of the variables included in the modified ratios may be based on the enterprise information sources presented in Table 1.

The estimation procedures presented in Table 1 are indicative and may vary depending on the accounting organization, production technology and information systems of the enterprise. In most industrial enterprises, however, the required information can be obtained from existing accounting records, enterprise resource

planning systems, production management systems, quality management systems and human resource databases without introducing additional accounting procedures.

Table 1. Typical estimation of the variables reflecting the impact of artificial intelligence

Variable	Typical calculation	Main enterprise data source
AI_R	Actual sales revenue after AI implementation − Estimated sales revenue without AI implementation	ERP sales module, CRM system, financial statements
AI_C	Total operating expenditures for software licenses, cloud services, maintenance, technical support and operation of AI systems	General ledger, ERP accounting system
AI_S	Baseline operating costs − Actual operating costs after AI implementation	Cost accounting system, ERP reports
AI_E	Energy costs before AI implementation − Energy costs after AI implementation	Energy management system, utility records
AI_Q	Reduction in scrap, rework, defective production and warranty costs	Quality management system, production records
AI_P	Production-related expenditures for implementing and operating AI technologies	Production accounting records
AI_L	Personnel costs before AI implementation − Personnel costs after AI implementation (including salaries and employer social security contributions)	Payroll accounting system, HR database
AI_T	Expenditures for personnel training, retraining and qualification related to AI implementation	HR information system, accounting records
AI_I	Investments in AI software, industrial robots, machine vision, digital twins, other intelligent assets	Fixed asset register, accounting records

An important advantage of the proposed conceptual framework is that all additional variables are expressed in monetary units, which ensures dimensional consistency of the modified efficiency ratios and facilitates their practical application in enterprise financial analysis. At the same time, the proposed variables can be estimated using accounting information that is already available in most industrial enterprises, making the framework suitable for practical implementation after appropriate empirical validation.

Furthermore, several limitations of the proposed framework should also be taken into consideration. The estimation of variables such as additional revenues AI_R, and cost savings AI_S attributable exclusively to AI may involve a certain degree of uncertainty because these effects are often influenced simultaneously by technological, organizational and market factors. Furthermore, the relative importance of the proposed variables may differ across industrial sectors depending on the production technology, enterprise size, and level of automation and maturity of artificial intelligence implementation. Therefore, future empirical studies should validate the proposed modified ratios using accounting and production data from industrial enterprises operating in different sectors and identify possible industry-specific adaptations of the proposed methodological framework.

The idea is for future research to be directed towards empirical testing of the proposed modified ratios under real conditions in industrial enterprises, as well as towards developing integral models for a comprehensive assessment of the impact of AI on the economic efficiency, competitiveness and sustainable development of industrial enterprises. The practical applicability of the proposed conceptual methodological framework is further illustrated through two hypothetical numerical examples presented in the following section.

5. Numerical Applications of the Proposed Modified Efficiency Ratios

The following hypothetical numerical examples demonstrate the practical application of the proposed modified efficiency ratios. The values are illustrative and are derived from typical accounting and management information maintained by industrial enterprises.

5.1. Example 1 - Calculation of the Modified Efficiency Ratio of Total Costs for AI Impact Assessment

Let's assume that the industrial enterprise reports *Net sales revenue* of EUR 12000000 and *Total costs* of EUR 9000000 during the reporting period then the classical *Efficiency ratio of total costs,* ER_{TC} is calculated as follows (1):

$$ER_{TC} = \frac{Net\ sales\ revenue}{Total\ costs} = \frac{12000000}{9000000} = 1{,}333.$$

Now, let's assume that the AI implementation generats *Additional sales revenue*, AI_R amounting to EUR 450000, while the *Operating costs for implementing and maintaining AI*, AI_C are EUR 180000. At the same time, the enterprise achieved the *Total cost savings due to AI implementation*, AI_S in the amount of EUR 320000. The *Modified efficiency ratio of total costs for AI impact assessment*, ER_{TC}^{AI} is calculated as (5):

$$ER_{TC}^{AI} = \frac{Net\ sales\ revenue + AI_R}{Total\ costs + AI_C - AI_S} = \frac{12000000 + 450000}{9000000 + 180000 - 320000} = 1{,}405.$$

The values of the variables are assumed to be obtained from the enterprise information systems as follows:

Net sales revenue – Statement of Comprehensive Income (Profit and Loss account) and ERP sales module;

Total costs – Statement of Comprehensive Income and General Ledger;

AI_R - comparison between actual sales revenue after AI implementation and the forecasted sales revenue without AI, based on ERP sales reports and CRM analyses;

AI_C - accounting records for software licences, cloud services, maintenance contracts, AI support and operation;

AI_S - cost accounting reports comparing operating costs before and after AI implementation.

The numerical example demonstrates that incorporating the economic effects of artificial intelligence into the proposed modified ratio (5) leads to a more comprehensive assessment of enterprise economic efficiency. In this illustrative case, the value of the the modified ratio (1,405) exceeds the value of the classical ratio (1,333) because the positive effects of additional revenues and operating cost savings outweigh the expenditures associated with the implementation and operation of artificial intelligence.

5.2. Example 2 - Calculation of the Modified Personnel Efficiency Ratio for AI Impact Assessment

Let's assume that the enterprise reports *Net sales revenue* of EUR 12000000 and *Personnel costs* of EUR 3600000 during the reporting period then the classical *Modified Personnel Efficiency Ratio,* ER_{AP} is calculated as follows (3):

$$ER_{AP} = \frac{Net\ sales\ revenue}{Personnel\ costs} = \frac{12000000}{3600000} = 3{,}333. \quad (3)$$

Now, let's assume that the AI implementation generats *Additional sales revenue*, AI_R of EUR 450000, *Personnel cost savings,* AI_L of EUR 280000 and *Costs of training personnel to work with AI systems*, AI_T of EUR 95000. The *Modified Personnel Efficiency Ratio for AI impact assessment,* ER_{AP}^{AI} is calculated as (7):

$$ER_{AP}^{AI} = \frac{Net\ sales\ revenue + AI_R}{Personnel\ costs - AI_L + AL_T} = \frac{12000000 + 450000}{3600000 - 280000 + 95000} = 3{,}646. \quad (7)$$

The values of the variables are assumed to be obtained from the enterprise information systems as follows:

Net sales revenue – Statement of Comprehensive Income (Profit and Loss account) and ERP sales module;

Personnel costs – payroll accounting records, payroll register and Statement of Comprehensive Income;

AI_L - comparison between personnel costs before and after AI implementation based on payroll accounting records, adjusted for changes in production volume;

AI_T – accounting records for personnel training, qualification programmes, seminars and certification related to artificial intelligence.

In this illustrative case, the modified personnel efficiency ratio (3.646) is higher than the classical personnel efficiency ratio (3.333), indicating that the personnel cost savings generated by artificial intelligence exceed the additional costs for personnel training. Consequently, the proposed ratio provides a more comprehensive assessment of personnel efficiency by taking into account both the reduction of personnel-related expenditures and the costs required for developing the necessary competencies for working with artificial intelligence systems.

6. Conclusion

Artificial intelligence is establishing itself as a key factor in increasing the economic efficiency of modern industrial enterprises. Its implementation impacts the main economic indicators by increasing revenues, reducing costs, increasing productivity, optimizing capital utilization, and improving production quality

This study proposes modified ratios for assessing economic efficiency, which expand the classic ratios by including specific additional variables that take into account the influence of artificial intelligence. The proposed modified ratios preserve the economic logic of the classical ratios, while simultaneously expanding their analytical capabilities in the context of Industry 4.0 and Industry 5.0. This allows for a more

complete and adequate measurement of the economic results of the implementation of AI technologies in industrial enterprises. The proposed modified ratios can be used as an analytical tool to assess the effectiveness of investments in artificial intelligence and to support management decisions related to the digital development of enterprises.

Acknowledgment

This study is financially supported by the Science Fund of the University of National and World Economy, Bulgaria under the project "Application of Artificial Intelligence for Sustainable Business Development in the Context of Climate Change" (Grant No. NID NI 35/2025/A).

References

[1] Singla A, Sukharevsky, Yee L, et al. The State of AI 2025: Agents, innovation, and transformation. Global Survey Report. McKinsey Global Institute. 2025 November, https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai#/.

[2] Sardi A, Modarelli G. Integrating artificial intelligence into performance measurement and management: a systematic literature review. International Journal of Productivity and Performance Management, 2026;75(6):1979–2022, doi: 10.1108/IJPPM-05-2025-0362

[3] Enholm IM, Papagiannidis E, Mikalef P, Krogstieet J. Artificial Intelligence and Business Value: a Literature Review. Inf. Syst. Front. 2022;24:1709-1734, doi:10.1007/s10796-021-10186-w.

[4] Czarnitzki D, Fernández GP, Rammer C. Artificial intelligence and firm-level productivity. Journal of Economic Behavior & Organization, 2023;211:188–205, doi: 10.1016/j.jebo.2023.05.008.

[5] Perifanis, N.-A., & Kitsios, F. Investigating the Influence of Artificial Intelligence on Business Value in the Digital Era of Strategy: A Literature Review. Information. 2023;14(2):1-42, art. no. 85, doi: 10.3390/info14020085.

[6] Xu X, Lu Y, Vogel-Heuser B, Wang L. Industry 4.0 and Industry 5.0—Inception, Conception and Perception. Journal of Manufacturing Systems. 2021;61:530-535, doi: 10.1016/j.jmsy.2021.10.006

[7] Horner H. The Shift from Industry 4.0 to 5.0: A Global Perspective on the Future of Manufacturing. Engineering Institute of Technology (EIT). 2025 April 11, https://www.eit.edu.au/industry-4-0-to-5-0-perspective-on-manufacturing/

[8] Islam MT, Sepanloo K, Woo S, Woo SH, Son Y-J. A Review of the Industry 4.0 to 5.0 Transition: Exploring the Intersection, Challenges, and Opportunities of Technology and Human–Machine Collaboration. Machines. 2025;13(4):1-34, art.no. 267, doi: 10.3390/machines13040267

[9] Tu Z, Chen Y, Liao Y. Artificial intelligence industrial policy and enterprise digital transformation. Finance Research Letters. 2026;91, art.no. 109384, doi:10.1016/j.frl.2025.109384

[10] Filippucci, F. et al. The impact of Artificial Intelligence on productivity, distribution and growth: Key mechanisms, initial evidence and policy challenges. OECD Artificial Intelligence Papers. 2024, no. 15, OECD Publishing, Paris, doi: 10.1787/8d900037-en.

[11] Amala KJ, et al. Artificial Intelligence for Industrial Efficiency. Empowering Sustainable Business Education and Training through AI. edited by E. Ivette Cota-Rivera, et al., IGI Global Scientific Publishing, 2026, pp. 141-166, doi: 10.4018/979-8-3373-5946-5.ch006

[12] Bobur N. Economic Efficiency of Widespread Use of Artificial Intelligence in Industrial Enterprises in the Context of the Uzbek Economy. Int. J. of Political Science and Economics, 2026;5(1):736-741, https://worldsiencepub.com/index.php/shokhal/article/view/6072

[13] Almulla D, Abbas M, Al-Alawi A, Alkooheji L. Process and Impact Evaluation of Artificial Intelligence in Managerial Accounting: A Systematic Literature Review. International Journal of Computing and Digital Systems. 2024;15(1):1467-1482, doi: 10.12785/ijcds/1501104

[14] Mikalef P, Gupta M. Artificial intelligence capability: Conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance. Information & Management. 2021;58(3):1-20, art. no. 103434, doi: 10.1016/j.im.2021.103434

[15] Hossain M A, Agnihotri R, Rushan MRI, Rahman MS, Sumi SF. Marketing analytics capability, artificial intelligence adoption, and firms' competitive advantage: Evidence from the manufacturing industry. Industrial Marketing Management. 2022;106:240–255, doi: 10.1016/j.indmarman.2022.08.017

[16] Abbas K. Management accounting and artificial intelligence: A comprehensive literature review and recommendations for future research. The British Accounting Review. 2026;58:1-26, art. no 101551, doi: 10.1016/j.bar.2025.101551
[17] Haefner N., Parida V, Gassmann O, Wincent J. Implementing and scaling artificial intelligence: A review, framework, and research agenda. Technological Forecasting & Social Change. 2023;197:1-11, art.no. 122878, doi: 10.1016/j.techfore.2023.122878
[18] Cosa M, Torelli R. Digital Transformation and Flexible Performance Management: A Systematic Literature Review of the Evolution of Performance Measurement Systems. Glob. J. Flex. Syst. Manag. 2024;25:445-466, doi: 10.1007/s40171-024-00409-9
[19] Berk J, DeMarzo P. Corporate Finance: The Core. 5th edition, Pearson, 2020.
[20] Brealey R, Myers St, Marcus A. Fundamentals of Corporate Finance, 10th edition, McGraw-Hill Education, 2020.
[21] Ross S, Westerfield R, Jaffe J, Jordan B. Corporate Finance, 13th edition, McGraw-Hill Education, 2022.
[22] Filippucci F, Gal P, Jona-Lasinio C, Leandro A, Nicoletti G. The Impact of Artificial Intelligence on Productivity, Distribution and Growth: Key mechanisms, initial evidence and policy challenges. OECD Artificial Intelligence Papers. 2024 April;15:1-63, OECD Publishing, Paris, doi: 10.1787/8d900037-en
[23] Brynjolfsson E, Li D, Raymond LR. Generative AI at Work. The Quarterly Journal of Economics. 2025May;140(2):889–942, doi: 10.1093/qje/qjae044
[24] Maslej N, Fattorini L, Perrault R, et al. Artificial Intelligence Index Report 2025. Stanford University - Institute for Human-Centered Artificial Intelligence. pp.1-456, 2024, doi: 10.48550/arXiv.2504.07139
[25] Chui M, Hazan E, Roberts R, et al. The Economic Potential of Generative AI: The Next Productivity Frontier. McKinsey & Company. pp. 1-63, June 2023, https://www.mckinsey.com/capabilities/tech-and-ai/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier
[26] Cao Z, Li M, Pavlou PA. AI in business research. Decision Sciences. 2024;55(6):518–532, doi: 10.1111/deci.12655
[27] Guettoufi Y. Assessing the impact of artificial intelligence adoption on corporate financial performance: Evidence from emerging markets. International Review of Economics & Finance. 2026;108:1-10, art. no. 105272, doi: 10.1016/j.iref.2026.105272.
[28] Bick A, Blandin A, Deming D, Fuchs-Schündeln N, Jessen J. Measuring AI Adoption among Firms: How You Ask Matters. St. Louis Fed on the Economy, June 2026, https://www.stlouisfed.org/on-the-economy/2026/jun/measuring-ai-adoption-firms-how-you-ask-matters

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)

© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261133

Digitalization and the International Competitiveness of Cultural Products: Evidence from China's Major Trading Partners

Yingyi Cao[a1] and Maoguang Pang[a]
[a] *School of Economics, Northwest Minzu University, China*

Abstract. Based on cultural trade and digital economy theories, this paper examines how digitalization affects the international competitiveness of cultural products. Using panel data of China's 39 major trading economies from 2008 to 2022, this study adopts principal component analysis and the entropy weight method to measure national digitalization levels and the comprehensive competitiveness of cultural products. Applying a two-way fixed‐effects model, together with instrumental variable estimation and propensity score matching to address endogeneity and ensure robustness, the empirical results confirm that digitalization significantly improves cultural product competitiveness. This enabling effect is stronger in developing countries and differs across cultural product categories. Mechanism tests further indicate that digitalization enhances competitiveness by lowering trade costs and promoting technological innovation, offering theoretical support and empirical evidence for the high-quality development of cultural trade.

Keywords. Digitalization; Cultural products; International competitiveness; Trade costs; Technological innovation

1. Introduction

Cultural trade is a core component of building China's cultural soft power and an essential part of its modern foreign trade system. To boost high-quality cultural trade development, Chinese authorities have issued targeted policies to expand trade scale, optimize industrial structure, and strengthen the global competitiveness of domestic cultural products. In the digital economy era, digital technologies have become a core driver of cultural industry upgrading, making digital-enabled cultural trade evolution a vital academic research topic [1].

However, the current literature exhibits two major limitations. First, relevant studies adopt fragmented theoretical perspectives and rarely integrate digital trade theory, cultural economics, international competitiveness theory, and innovation diffusion theory into a unified analytical system. Traditional static trade frameworks are insufficient to explain the dynamic empowering effects of digitalization. Second, existing theoretical explanations remain predominantly descriptive. Few studies construct a rigorous mediating mechanism model to systematically clarify how

[1] Corresponding Author, Yingyi Cao, Email: 79059732@qq.com

digitalization reshapes cultural product competitiveness through cost optimization and innovative development, resulting in insufficient theoretical support for empirical hypotheses.

This creates a critical research gap. Unlike fixed traditional trade factors, digitalization dynamically restructures production factors, optimizes market matching, and renovates cross-border transaction mechanisms. Traditional theories tailored for static comparative advantages cannot adapt to the new competitive logic of digital cultural trade, highlighting the necessity of an updated framework integrating digital economic mechanisms.

Against this backdrop, this paper employs 2008–2022 panel data of 39 major Chinese trading partners to explore digitalization's impact on cultural product competitiveness. Theoretically, it constructs an integrated framework of digital economy and cultural trade, clarifying two core transmission paths: trade cost reduction and technological innovation, which remedies the static theoretical bias of prior studies. Empirically, differing from research on traditional factors or general commodity trade, this paper targets cultural products to systematically examine digitalization's baseline effect and heterogeneous impacts. This study enriches cultural trade theories and provides practical references for advancing digital-cultural integration and improving the global competitiveness of China's cultural products.

2. Literature Review and Theoretical Framework

2.1 Literature Review and Theoretical Foundation

Research on the international competitiveness of cultural products is traditionally rooted in classical trade theory and cultural economics. Existing studies predominantly explain cross-border cultural trade performance relying on static exogenous factors, including geographical distance, cultural proximity, economic scale and institutional quality [2–4]. Such theoretical paradigms take static factor endowments as the core analytical logic, assuming that cultural trade advantages are relatively stable and determined by inherent national conditions. With the proliferation of digital technologies, recent studies have verified that digitalization substantially reshapes cultural trade patterns by breaking temporal and spatial trade constraints, optimizing industrial operation modes, and lowering cross-border trade frictions [5–7].

Nevertheless, the current literature has two core theoretical deficiencies, which this study aims to address. First, prior research adopts scattered theoretical perspectives and lacks an integrated analytical system that unifies digital trade theory, cultural economics, international competitiveness theory, and innovation diffusion theory. Traditional static trade frameworks cannot explain the dynamic, technology-endogenous evolution of contemporary digital cultural trade, resulting in obvious theoretical adaptation defects. Second, existing mechanism explanations are mostly phenomenon-based and descriptive, lacking rigorous theoretical deduction and complete mediating logical chains. Few studies systematically clarify how digitalization converts technological advantages into cultural product competitive advantages through intermediate paths. To fill this gap, this paper constructs a standardized multi-theory theoretical transmission model: Digitalization → Reduction of cross-border transaction costs / Improvement of industrial innovation capability → Enhancement of cultural product international competitiveness.

Specifically, digital trade theory undertakes the theoretical interpretation of digital transaction efficiency and cost-saving effects; innovation diffusion theory explains the iterative upgrading of cultural industry technology and product innovation; cultural economics clarifies the supply-demand matching optimization of cultural products with dual economic and cultural attributes; international competitiveness theory further solidifies the logical conversion of cost advantages and innovation advantages into global market competitiveness. This multi-dimensional theoretical integration makes the mechanism analysis deductive and theoretically grounded, rather than simple empirical description.

2.2 Digitalization and Cultural Product Competitiveness

Based on the integrated theoretical framework above, digitalization exerts a comprehensive positive impact on the international competitiveness of cultural products by optimizing industrial factor allocation, driving digital transformation, and realizing precise global market matching. From the production perspective, in line with innovation diffusion theory, the widespread penetration of digital technologies reconstructs the traditional manual and extensive production mode of cultural industries, effectively improving production efficiency and product refinement, and supporting diversified and personalized cultural product supply to adapt to heterogeneous global consumer demands [8]. From the market and transaction perspective, consistent with digital trade theory, digitalization eliminates the time-space limitations of traditional cultural trade [9]. It enables cultural enterprises to efficiently identify overseas potential markets and capture global niche resources [10], accelerates enterprise digital transformation, improves the matching degree between cultural product supply and international market demand, expands the coverage and depth of cross-border cultural marketing, and significantly optimizes cross-border transaction efficiency [11]. Different from descriptive discussions in previous studies, this study further confirms that digitalization forms sustainable competitive advantages for cultural products through systematic technological empowerment and market optimization. Synthesizing the above theoretical deduction and existing empirical evidence, this paper proposes the core baseline hypothesis:

Hypothesis 1. The improvement of digitalization level significantly enhances the international competitiveness of cultural products.

2.3 Digitalization, Trade Cost Reduction and Competitiveness

Combined with digital trade theory, this study clarifies the cost-reduction transmission mechanism of digitalization empowering cultural product competitiveness. Traditional cross-border cultural trade faces high endogenous transaction costs caused by information asymmetry, cumbersome offline procedures and institutional trade barriers, which restrict the global market expansion of cultural products. According to international competitiveness theory, the reduction of trade costs helps expand cultural product export scale and improve overseas market penetration rate, forming stable price and market competitive advantages [12]. Digital transformation fundamentally optimizes cross-border trade procedures, breaks time-space constraints and procedural frictions, and greatly improves the convenience and efficiency of cultural product cross-border transactions [13]. Digital trade empowerment accelerates the cross-border free flow of information, technology and human capital, effectively reducing multi-dimensional

transaction costs including overseas market search costs, cross-border communication costs and institutional compliance costs [14]. This study further upgrades the previous descriptive mechanism analysis to theoretical deduction based on digital trade and competitiveness theories, and proposes the first mediating hypothesis:

Hypothesis 2. Digitalization improves the international competitiveness of cultural products by reducing cross-border trade costs.

2.4 Digitalization, Technological Innovation and Competitiveness

Based on innovation diffusion theory and cultural economics, this paper constructs the innovation-driven internal mechanism of digitalization empowering cultural trade competitiveness. Classic cultural industrial theory points out that the integration of cultural production and modern technology is the core driving force of industrial upgrading and value appreciation [15]. Digital technology, as a universal innovative technology, has strong industrial penetration and diffusion characteristics. It can realize the comprehensive innovation and iteration of cultural product creation, production and dissemination links, drive the transformation and upgrading of traditional cultural industries, and cultivate new forms and new advantages of modern cultural trade [16]. The continuous diffusion of digital innovation activates the innovation vitality of the cultural industry, optimizes the quality and creative connotation of cultural products, and forms differentiated non-price competitive advantages in the global market [17]. Combined with international competitiveness theory, industrial technological innovation and product upgrading can effectively consolidate the long-term international competitive capacity of cultural products. On the basis of standardized theoretical deduction, this paper proposes the second mediating hypothesis:

Hypothesis 3. Digitalization enhances the international competitiveness of cultural products by promoting technological innovation and innovation diffusion in the cultural industry.

3. Research Design

3.1. Model Specification

Based on the theoretical analysis, to examine the impact of the level of digitalization on the international competitiveness of cultural products, this paper constructs the following baseline regression model.

$$IC_{it} = \beta_0 + \beta_1 DL_{it} + \lambda Control_{it} + \mu_i + \delta_t + \varepsilon_{it} \quad (1)$$

where the subscripts i and t denote country and year, respectively. IC_{it} is the dependent variable, representing the international competitiveness of cultural products of country i in year t. DL_{it} is the core explanatory variable, representing the digitalization level of country i in year t. $Control_{it}$ denotes a set of control variables, μ_i denotes country fixed effects, absorbing all time-invariant national characteristics such as cultural endowments, institutional quality and geographical conditions. δ_t denotes year fixed effects, capturing global time-varying shocks including economic fluctuations and universal technological progress and ε_{it} is the error term.

3.2. Variable Description

Dependent variable: international competitiveness of cultural products (*IC*). Referring to UNESCO's definition of cultural products and the 2009 Framework for Cultural Statistics, this paper selects core cultural products as the research object. Drawing on Qi and He [18], the entropy weight method is used to calculate the weights of each indicator and obtain the comprehensive score of the international competitiveness of cultural products for each country-year.

Core explanatory variable: digitalization level (*DL*). Referring to the ICT Development Index released by the International Telecommunication Union and drawing on the indicator systems constructed by Lin et al. [19] and Duan and Feng [20], this paper establishes the indicator system shown in Table 1 and uses principal component analysis to measure the level of digitalization.

Table 1. Indicator System of the Digitalization Level

Level indicator	Secondary indicator	Tertiary indicator	Unit
Digitalization level (DL)	Digital infrastructure	Fixed telephone subscriptions x1	Per 100 people
		Mobile-cellular subscriptions x2	Per 100 people
		Fixed broadband subscriptions x3	Per 100 people
		Individuals using the Internet x4	%
		Secure Internet servers x5	Per million people
	Digital trade development	ICT goods exports x6	%
		ICT goods imports x7	%
		High-technology exports x8	%
		R&D researchers x9	Per million people
	Digital application capacity	Scientific and technical journal articles x10	Articles
		R&D expenditure as a percentage of GDP x11	%
		Resident patent applications x12	Applications

Control Variables: This paper selects five country-level control variables that affect cultural trade competitiveness: (1) Economic development level (lnGDP), measured by per capita GDP; (2) Population size (lnPOP), reflecting market scale; (3) Foreign direct investment (FDI), measured by the ratio of FDI net inflow to GDP; (4) Resident consumption level (CPI), reflecting domestic consumption capacity; (5) Government expenditure (GE), measured by the ratio of government fiscal expenditure to GDP.

3.3. Data Sources

This paper selects unbalanced panel data of 39 major Chinese trading partners from 2008 to 2022. The sample countries cover developed and developing economies across Asia, Europe, America and Africa, accounting for more than 83% of global cultural product trade volume, with strong representativeness. All original data are derived from the World Bank WDI Database and UN Comtrade Database, and individual missing values are supplemented by linear interpolation to ensure data integrity.

4. Empirical Tests

4.1. Baseline Regression Results

Table 2 reports the baseline regression results of the two-way fixed-effects model. Column (1) presents the regression result without control variables, and columns (2) to (6) gradually add control variables. The results show that the coefficient of digitalization level is consistently significantly positive at the 1% level. After controlling all time-invariant national individual characteristics and global time-varying shocks, digitalization is positively associated with the international competitiveness of cultural products. This finding supports Hypothesis 1, indicating that the improvement of digitalization level helps strengthen cultural product competitiveness.

Table 2. Baseline Regression Results

Variable	**(1)** IC	**(2)** IC	**(3)** IC	**(4)** IC	**(5)** IC	**(6)** IC
DL	0.113***	0.259***	0.174***	0.168***	0.167***	0.148***
	(0.028)	(0.037)	(0.026)	(0.026)	(0.026)	(0.027)
lnGDP		-0.037***	0.018***	0.019***	0.019***	0.025***
		(0.007)	(0.005)	(0.005)	(0.006)	(0.006)
lnPOP			0.056***	0.059***	0.057***	0.057***
			(0.004)	(0.004)	(0.004)	(0.004)
FDI				0.065	0.066	0.049
				(0.046)	(0.046)	(0.043)
CPI					-0.004	-0.003
					(0.020)	(0.020)
GE						-0.032***
						(0.012)
Constant	0.075***	0.383***	-1.099***	-1.121***	-1.115***	-1.070***
	(0.020)	(0.064)	(0.108)	(0.109)	(0.112)	(0.112)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
N	585	585	585	585	585	585
R-squared	0.031	0.106	0.436	0.437	0.437	0.447

Note. ***, ** and * denote significance at the 1%, 5% and 10% levels, respectively. The same applies to the following tables.

4.2. Endogeneity Treatment

4.2.1. Reverse Causality

To alleviate reverse causality and omitted variable bias, this paper adopts two instrumental variable (IV) strategies with rigorous exclusion restriction justification. First, this paper adopts one-period lagged digitalization level as a supplementary instrumental variable, with sufficient theoretical justification: lagged digitalization affects current cultural competitiveness only through current digital development, and the time-lagged impact of digital technology diffusion conforms to industrial economic laws, satisfying the exclusion restriction. [21]

Second, referring to classic panel IV design, this paper constructs historical digital infrastructure instrumental variable: the interaction term of 2004 fixed telephone penetration (historical exogenous infrastructure endowment, not affected by subsequent

cultural trade performance) and annual global digital growth shock. This historical instrumental variable satisfies exogeneity (pre-determined historical data) and relevance (strong correlation with current digitalization level).[22][23][24]

The IV test results in Table 3 show that both instrumental variables pass the under-identification and weak instrumental variable tests (Cragg-Donald Wald F and Kleibergen-Paap rk F statistics far exceed critical values). The second-stage regression coefficients remain significantly positive, confirming that the positive correlation between digitalization and cultural product competitiveness is robust after correcting endogeneity bias.

Table 3. Endogeneity Test Results

Variable	**(1)** IV method First stage DL	**(2)** IV method Second stage IC	**(3)** IV method First stage DL	**(4)** IV method Second stage IC	**(5)** PSM method PSM method IC
IV1	0.959***				
	(43.33)				
IV2			0.0123***		
			(24.00)		
DL		0.150***		0.2190***	0.177***
		(5.20)		(6.11)	(3.16)
Constant	0.009	-1.069***	0.8134***	-1.0108***	-1.139***
	(0.16)	(-9.17)	(5.62)	(-8.33)	(-6.73)
Control variables	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes	Yes
N	546	546	585	585	293
R-squared	0.850	0.446	0.470	0.440	0.455
Cragg-Donald Wald F	2984.070		501.030		
K-P rk Wald F	1877.340		575.940		
K-P rk LM	181.978***		143.778***		

Note. ***, ** and * denote significance at the 1%, 5% and 10% levels, respectively. The same applies to the following tables.

4.2.2. Selection Bias

To eliminate self-selection bias, this paper adopts the propensity score matching (PSM) method. [25] [26] Samples are divided into treatment group (above median digitalization level) and control group (below median level). One-to-one nearest neighbor matching, radius matching and kernel matching are adopted for robustness. The balance test shows that the standardized bias of all covariates is controlled below 5% after matching, and the propensity score distribution of the two groups is highly consistent, indicating effective matching. The PSM regression result is significantly positive, further verifying the core conclusion.

4.3. Robustness Check

This paper conducts multiple robustness tests including lagged explanatory variable treatment, dependent variable replacement, and sample period adjustment. All regression results remain consistent with the baseline conclusion, proving the reliability of the empirical findings. We follow approach by applying first- and second-order lags to the explanatory variables [27]. Drawing on prior research, we adjust the sample period by

excluding data from 2008 and 2020—years marked by exceptional circumstances—to avoid bias [28]. Detailed results of these robustness tests are shown in Table 4.

Table 4. The results of the robustness test

Variable	**(1)** IC One period behind schedule	**(2)** IC Lagging by two periods	**(3)** IC_1 Change the dependent variable	**(4)** IC_2 Change the dependent variable	**(5)** IC Adjust the sample period
DL			0.126***	0.230***	0.149***
			(5.707)	(9.117)	(5.036)
L.DL	0.144***				
	(5.116)				
L2.DL		0.143***			
		(4.964)			
Constant			-0.861***	0.026	-1.089***
			(-9.464)	(0.270)	(-8.822)
Control variables	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes	Yes
N	546	507	585	585	507
R-squared	0.441	0.437	0.449	0.664	0.445

Note. ***, ** and * denote significance at the 1%, 5% and 10% levels, respectively. The same applies to the following tables.

4.4. Heterogeneity Test

4.4.1. Economic Development Heterogeneity

Referring to the World Bank national income classification standard, this paper divides sample countries into developed economies and developing economies to conduct heterogeneity analysis. The interaction term coefficient is significantly negative, indicating that the positive association between digitalization and cultural product competitiveness is more prominent in developing countries. This may be because developing countries have greater room for digital transformation, and digital empowerment can more effectively break the original static endowment constraints and release cultural trade dividends. This paper will follow the approach of Li (2019) [29]. The test results are shown in Table 5.

4.4.2. Product Type Heterogeneity

Based on UNESCO official cultural product classification standards, this paper divides cultural products into six categories for segmented regression. This article refers to research [30], calculates the international competitiveness of each type of product and separately examines the impact of digitalization levels on each category. The results are shown in Table 6.

The results show that digitalization has significant positive associations with the competitiveness of performance and celebration products, visual arts and handicrafts, books and newspapers, and design and creative services.

Table 5. Results of the test for heterogeneity in economic development levels

Variable	(1)	(2)
	IC	IC
DL	0.148***	0.191***
	(0.027)	(4.780)
dev_DL		-0.059*
		(-1.897)
Constant	-1.070***	-1.173***
	(0.112)	(-8.592)
Control variables	Yes	Yes
Year fixed effects	Yes	Yes
N	585	585
R-squared	0.447	0.451

Note. ***, ** and * denote significance at the 1%, 5% and 10% levels, respectively. The same applies to the following tables.

Table 6. Test results for heterogeneity by product type

Variable	(1)	(1)	(2)	(3)	(4)	(6)
	IC_1	IC_2	IC_3	IC_4	IC_5	IC_6
	Class A	Class B	Class C	Class D	Class E	Class F
DL	0.016	0.106***	0.028***	0.032***	0.013	0.060***
	(0.859)	(13.174)	(2.901)	(4.392)	(0.810)	(3.905)
Constant	-0.694***	-0.186***	-0.447***	-0.462***	-0.362***	-0.408***
	(-11.417)	(-6.899)	(-13.928)	(-19.101)	(-6.992)	(-7.909)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
N	546	585	585	585	585	585
R-squared	0.263	0.463	0.444	0.504	0.169	0.190

Note. ***, ** and * denote significance at the 1%, 5% and 10% levels, respectively. The same applies to the following tables.

These categories are highly dependent on creation, marketing and cross-border dissemination, which are highly sensitive to digital empowerment. In contrast, cultural heritage and audiovisual products have relatively insignificant effects, which may be restricted by cultural attribute specificity and industry regulation.

5. Mechanism Verification

5.1. Model Specification

To verify the mechanism by which the level of digitalization affects the international competitiveness of cultural products, that is, whether the level of digitalization can positively promote the international competitiveness of cultural products through trade cost savings and technological innovation. This paper will refer to two-step method [31] and, based on the baseline model, construct the following model:

$$Cost_{it} = \beta_0 + \beta_1 DL_{it} + \gamma Control_{it} + \mu_i + \delta_t + \varepsilon_{it} \quad (2)$$

$$Innov_{it} = \beta_0 + \beta_1 DL_{it} + \gamma Control_{it} + \mu_i + \delta_t + \varepsilon_{it} \quad (3)$$

Among them, the subscripts i and t respectively represent the country and the year. $Cost_{it}$ represents the dependent variable, indicating the trade cost of country *i* in year *t*. $Innov_{it}$ represents the dependent variable, signifying the innovation capability of country i in year t. DL_{it} serves as the core explanatory variable, indicating the digitalization level of country i in year t. Control denotes a set of control variables, and ε_{it} represents the error term.

5.2. Mechanism Variable

For the trade cost variable (*Cost*), This article selects the Global Economic Freedom Index (Freedom) as the indicator to evaluate the trade cost of economies [32]. The Global Economic Freedom Index is compiled and released by the Heritage Foundation of the United States, and this index assesses from multiple dimensions such as rule of law, efficiency, and market. The data is sourced from the website of the Heritage Foundation.

For the technological innovation variable (*Innov*), This article refers to the measurement of innovation capacity, and selects the Global Innovation Index (GII) to measure the innovation capacity of various countries [33]. The data is sourced from the World Intellectual Property Organization, and missing data is supplemented using interpolation methods.

5.3. Test Results

Table 7 reports the empirical results of the dual mediating mechanisms. Column (1) is the baseline regression result consistent with the previous analysis, verifying the total positive effect of digitalization on cultural product competitiveness. Column (2) takes trade cost as the explained variable; the regression coefficient of DL is significantly negative at the 1% level, indicating that the improvement of digitalization level can effectively reduce cross-border trade costs of cultural products. Column (3) takes technological innovation as the explained variable; the coefficient of DL is significantly positive, confirming that digital technology diffusion significantly promotes the technological innovation and industrial upgrading of cultural industries.

Column (4) incorporates both mediating variables and the core explanatory variable simultaneously. The results show that the coefficients of trade cost are significantly negative, and the coefficients of technological innovation are significantly positive. Meanwhile, the coefficient of digitalization level is still significantly positive but slightly lower than the baseline regression coefficient, which indicates that trade cost reduction

and technological innovation upgrading play significant partial mediating roles in the empowerment process of digitalization on cultural product competitiveness.

The above results fully verify Hypothesis 2 and Hypothesis 3, demonstrating the complete transmission chain of digitalization → trade cost reduction / technological innovation → enhanced international competitiveness of cultural products.

Table 7. Results of Mediating Mechanism Test

Variable	(1)	(2)	(3)	(4)
	lnCost	lnCost	Innov	Innov
DL	0.463***	0.109***	0.420***	0.227***
	(0.022)	(0.023)	(0.016)	(0.017)
lnGDP		0.078***		0.050***
		(0.004)		(0.004)
lnPOP		-0.019***		0.013***
		(0.003)		(0.002)
FDI		0.078***		0.042
		(0.028)		(0.033)
CPI		-0.027**		-0.037**
		(0.012)		(0.016)
lnGE	4.029***	4.080***		-0.006
	(0.018)	(0.086)		(0.007)
Control variables	No	Yes	No	Yes
Year fixed effects	Yes	Yes	Yes	Yes
N	585	585	585	585
R-squared	0.384	0.739	0.871	0.915
F	29.340	117.970	252.420	187.810

Note. ***, ** and * denote significance at the 1%, 5% and 10% levels, respectively. The same applies to the following tables.

5.4. Mechanism Discussion

Based on the above empirical results, this paper further clarifies the internal logic of the dual mediating paths. On the cost reduction path, digitalization breaks the time-space constraints and information asymmetry of traditional cultural trade. Digital information matching, online transaction systems and intelligent customs clearance procedures effectively reduce multi-dimensional transaction frictions including search cost, communication cost and institutional compliance cost. The reduction of trade costs lowers the threshold for cultural enterprises to enter overseas markets, expands the export scale and market penetration of cultural products, and ultimately transforms cost advantages into international competitive advantages.

On the technological innovation path, digital technology, as a universal innovative factor, penetrates and reconstructs the entire cultural industry chain covering creation, production, dissemination and operation. It drives the iterative upgrading of cultural product forms and creative content, stimulates the innovative vitality of the cultural industry, and forms differentiated non-price competitive advantages. Different from traditional static factor endowment advantages, digital innovation diffusion builds

endogenous and sustainable competitive capabilities for cultural products, which fundamentally improves the international market competitiveness of cultural products.

6. Discussion and Conclusion

6.1 Main Research Conclusions

Based on the 2008–2022 panel data of 39 major Chinese trading partners, this study integrates multiple theories to construct a digital empowerment framework for cultural product competitiveness and conducts rigorous empirical analyses. The results reveal that digitalization significantly and stably enhances the international competitiveness of cultural products by optimizing the whole industrial chain of cultural creation, cross-border transactions and overseas marketing, with the conclusion robust after endogeneity correction and multiple robustness tests. Obvious heterogeneity exists in the empowerment effect.

Digitalization presents a more prominent promotional effect in developing economies with greater digital transformation potential and marginal dividend space. In terms of product differences, it substantially improves the competitiveness of digitally sensitive categories including performing arts, visual crafts, publishing products and creative services, while exerting insignificant influences on heritage and audiovisual products constrained by inherent attributes and industry regulations. Mechanism tests confirm two valid transmission paths: digitalization reduces cross-border trade frictions to form cost advantages and facilitates industrial innovation diffusion to build differentiated non-price advantages, forming a complete dynamic empowerment mechanism for digital cultural trade.

6.2 Theoretical Contributions

This study makes three key theoretical marginal contributions to digital cultural trade research. First, it transcends the static factor endowment framework adopted in traditional studies. By incorporating dynamic digital technological factors and constructing a multi-theory integrated analytical system, it remedies the fragmented and static theoretical perspectives in existing literature. Second, this paper systematically identifies the dual transmission mechanism of digital empowerment, clarifying the logical relationships among digitalization, trade costs, technological innovation and cultural product competitiveness. It improves theoretical explanatory power by eliminating the superficial descriptive mechanism analysis of prior studies. Third, it applies standardized PCA and entropy weight methods for variable measurement, solving opaque criteria and poor reproducibility in previous research and providing a unified empirical paradigm for future studies.

6.3 Practical Policy Implications

This study offers targeted policy implications for advancing China's digital cultural trade and strengthening cultural product competitiveness. Governments should accelerate digital infrastructure construction and deepen the integration of digital technology and cultural industries to fully release trade development dividends. Differentiated industrial

strategies are essential: digital empowerment should prioritize performing arts, crafts and publishing products to build international cultural brands, while heritage and audiovisual sectors require moderate digital optimization under standardized supervision. Additionally, dual improvements in cost control and innovation are vital. Digital cross-border trade platforms can reduce corporate transaction costs, and increased digital R&D investment fosters product and business model innovation to form sustainable competitive advantages. Furthermore, differentiated international cooperation strategies should be implemented to expand markets in developing economies and absorb advanced digital operation and IP protection experience from developed economies.

6.4 Research Limitations and Future Prospects

This study has several limitations deserving further exploration. Firstly, it adopts national macro panel data for empirical analysis, failing to capture micro-level heterogeneity across cultural enterprises and individual products. Future research can integrate firm-level data to investigate digitalization's differential impacts on trade performance among enterprises with varied scales, ownerships and innovation capacities. Secondly, this paper only verifies the dual mediating effects of trade costs and technological innovation. It ignores potential moderating factors such as institutional environment, cultural distance and digital intellectual property protection, which can be incorporated to construct a more comprehensive theoretical framework in follow-up studies. Thirdly, constrained by data accessibility, this study uses annual national panel data. With improved statistical systems, future research can adopt higher-frequency and refined product-level data to further validate and extend the present findings.

References

[1] Jiang Xiaowei. Technology and Culture in the Digital Age [J]. Chinese Social Sciences, 2021, (08): 4-34 + 204.

[2] Luo Libin, Wei Changfeng, Wang Moxin. Cultural Distance, Income Level and Cultural Product Exports: A Case Study of Movies [J]. Cultural Industry Research, 2022, (02): 28-47.

[3] Wei Yonggui, Li Hong. Research on the Influencing Factors of China's Cultural Product Export Trade - Based on the Three-Dimensional Distances of Culture, Geography and Institution [J]. Modern Finance (Tianjin University of Finance and Economics Journal), 2016, 36(10): 103-113.

[4] Zhou Shengli, Lu Ronghui. Research on the International Competitiveness of China's Cultural Product Trade and Its Influencing Factors - Based on the Supply and Demand Perspective [J]. Price Monthly, 2019, (07): 51-59.

[5] Li Huailiang, Tong Xuena. New Trends in International Cultural Trade under Digitalization [J]. Chinese Culture Research, 2012, (03): 130-138.

[6] Fang Ying, Wu Xuechun. The Positive Effects and Promotion Strategies of Digitalization Development of China's Cultural Trade [J]. Modern Communication (Journal of China Media University), 2020, 42(11): 1-7.

[7] Huang Fu Tao. Research on High-Quality Digital Development of China's Cultural Trade [J]. Technology Economy and Management Research, 2021, (10): 119-122.

[8] Li Lei, Chang Shixian, Wang Ziqi. The Theoretical Logic, Realistic Problems and Governance Path of Cultural Digital Development in Western Regions [J]. Xi'an University of Finance and Economics Journal, 2024, 37(06): 13-25.

[9] Yu Qinming, Sun Guohui, Yan Chaosheng, et al. Research on the Cultivation of Digital Talents in Cultural Industry in the Digital Age [J]. Cultural Innovation Comparative Research, 2024, 8(35): 91-95.

[10] Li Wenting. A Brief Analysis of Precise Management of Enterprises under Digitalization Model [J]. China Management Informatization, 2021, 24(02): 119-122.

[11] Pan Ailing, Wang Xue. How Digital Transformation Promotes High-Quality Development of Cultural Enterprises [J]. Shenzhen University Journal (Humanities and Social Sciences Edition), 2023, 40(04): 44-54.
[12] Fang Hui, Zhang Xiaoye. Measurement of China's Cultural Digital Level and Its Export Effectiveness Research [J]. Shandong University Journal (Philosophy and Social Sciences Edition), 2022, (03): 38-51.
[13] Wen Dongwei, Xu Wenqing. Digital Innovation Technology and International Trade Reconfiguration: An Analytical Framework [J]. Xi'an Jiaotong University Journal (Social Sciences Edition), 2025, 45(01): 123-137.
[14] Liu Bin, Zou Tianhua, Zhang Han. The Impact of the "Belt and Road" Initiative on Cultural Product Exports - An Empirical Analysis Based on the Multi-Period DID Model [J]. Southern Economy, 2023, (07): 135-152.
[15] Yuan Zhihui, Xie Xianjun. Digitalization, R&D Cooperation and Enterprise Technological Innovation Quality [J]. Technology Economy and Management Research, 2025, (01): 8-14.
[16] Zhang Yajun, Xia Jiazhang. The Driving Mechanism, Challenges and Countermeasures of Cultural and Technology Integration [J]. Administrative Management Reform, 2024, (06): 60-69.
[17] Guo Ran, Yuan Yijun, Zhang Yongxin. Internet Development, Technological Innovation and International Competitiveness of Manufacturing - An Empirical Analysis Based on Cross-border Data [J]. Economic Issues Exploration, 2021, (01): 171-180.
[18] Qi Wei, He Aijuan. Measurement and International Comparison of China's Cultural Product Export Competitiveness [J]. Statistics and Decision, 2020, 36(05): 91-94.
[19] Lin Xianyan, Tang Hongbo, Liu Yali, et al. Research on the Information and Communication Technology/Digital Development Levels of BRICS Countries [J]. Science and Technology Management Research, 2023, 43(01): 146-155.
[20] Duan Dingyuan, Feng Zongxian. Research on the Impact of Digital Development Levels of "Belt and Road" Countries on Low-Carbon Green Performance [J]. Economic Issues Exploration, 2023, (05): 158-176.
[21] Xu Helian, Cheng Lihong, Sun Tianyang. The Enhancing Effect of Manufacturing Input Serviceization on Domestic Added Value of Enterprises' Exports - An Empirical Study Based on Micro Enterprises in China's Manufacturing Industry [J]. China Industrial Economy, 2017, (10): 62-80.
[22] Huang Qunhui, Yu Yongze, Zhang Songlin. Internet Development and the Improvement of Manufacturing Productivity: Internal Mechanism and Chinese Experience [J]. China Industrial Economy, 2019, (08): 5-23.
[23] Nunn, N, N Qian. US Food Aid and Civil Conflict [J]. The American Economic Review, 2014, 104(6): 1630-1666.
[24] Wang Yan, Sheng Xiaodan. Research on the Impact of Digital Economic Development of "Belt and Road" Countries on China's Cross-Border E-commerce Exports [J]. Modern Finance (Tianjin University Journal), 2024, 44(05): 22-44.
[25] Wang Jingru, Yao Yi. Enterprise Digital Transformation, Fault Tolerance Mechanism and R&D Innovation [J]. Foreign Economics and Management, 2023, 45(09): 38-53.
[26] Min Zhihui, Xiong Xin. Research on the Impact of Digital Transformation on Enterprise Financialization - Based on the Motivation Perspective [J]. Financial Management Research, 2024, (06): 118-125.
[27] Wang Bin. The Impact of Digital Economy on China's Carbon Emissions from the Perspective of Global Value Chain [J]. Resources Science, 2023, 45(09): 1899-1911.
[28] Zhang Tianding, Xie Yutong. Digital Transformation and the Upgrading of China's Manufacturing Enterprises' Global Value Chain - An Empirical Study Based on the Perspective of Digital Mergers and Acquisitions [J]. Asia-Pacific Economy, 2024, (02): 96-108.
[29] Li J, Lu Y, Song H, et al. Long-term Impact of Trade Liberalization on Human Capital Formation [J]. Journal of Comparative Economics, 2019, 47(4): 946-961.
[30] Zhao Yuhuan, Zuo Sulin. Research on the Impact of Digital Economic Development of Importing Countries (Regions) on China's Cultural Product Exports [J]. International Trade Issues, 2024, (01): 122-139.
[31] Jiang Tie. Mediating Effects and Moderating Effects in Causal Inference Empirical Research [J]. China Industrial Economy, 2022, (05): 100-120.
[32] Qu Ruxiao, Yang Xiu, Liu Yang. Cultural Differences, Trade Costs and China's Cultural Product Exports [J]. World Economy, 2015, 38(09): 130-143.
[33] Pu Xiaosong, Cao Peijie, Zhang Yongjun. The Relationship between Education Development Level and National Innovation Capacity - A Quantile Regression Analysis Based on International Data [J]. Journal of Dalian University of Technology (Social Science Edition), 2019, 40(04): 104-109.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)

© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261134

Breaking the "Last Mile" of Agricultural ESG: Evidence from 506 Smallholder Farmers

Jinfang Qian[a, b,c], Zixu Xia[a], Baijie Wang[a], and Fang Guo [a,1]
[a] *School of Management, Dalian Polytechnic University, China;*
[b] *School of Economics and Management, Dalian University of Technology, China;*
[c] *China Business Executives Academy, Dalian, China.*
ORCiD ID: Jinfang Qian https://orcid.org/ 0009-0009-4931-600X

Abstract. This study investigates how key forms of external support and prior behavioral experience shape smallholder farmers' willingness to participate in agricultural ESG practices. Using a mixed-methods design, we integrate grounded theory coding with econometric analysis of survey data from 506 smallholders in Hebei and Liaoning provinces. The qualitative analysis identifies the mechanism that technical guidance and government subsidy operate through two cognitive channels. The empirical results show that technical guidance influences willingness primarily by enhancing ESG value perception, while government subsidy operates both directly and through perceived risk alleviation. Past green behavior remains the strongest direct predictor of participation willingness. These findings shift attention from a broad list of determinants to differentiated cognitive mechanisms and provide actionable implications for technical services, subsidy design, and market-oriented incentives that can help bridge the agricultural ESG "last mile".

Keywords. Agricultural ESG; Smallholder farmers; Green production behavior; Mixed-methods research

1. Introduction

Agriculture constitutes the foundation of the national economy, and its sustainable development is integral to both food security and the broader ecological civilization agenda. ESG (Environmental, Social, and Governance), as an internationally recognized framework for measuring organizational sustainability performance, provides a systematic evaluation tool and actionable guideline for agricultural green transformation. The 2024 "No. 1 Central Document" designated "green agricultural development" as a key priority. However, China's agricultural sector is fundamentally characterized by "large country, small farmers," with smallholders account for over 98% of agricultural operators and manage approximately 70% of the total cultivated land area[1]. As the primary agents of agricultural production at the source of the supply chain, smallholders play a pivotal role in achieving green agricultural transformation and implementing ESG projects across the agricultural industry.

[1] Corresponding Author: Fang Guo, E-mail: guofang@dlpu.edu.cn.
This work was supported by the Liaoning Province Natural Science Foundation Project [grant number 2025-BS-0466]; the Liaoning Provincial Department of Education Fundamental Research Project for Universities [grant number LJ112610152013].

The so-called "last mile" dilemma of ESG refers to the systemic failure in measurement, incentive, and governance that arises when corporate-level ESG standards are transmitted to smallholder farmers. Smallholders have long occupied the status of a "silent majority" within the ESG discourse: their production units are dispersed, their resource scale is limited, and their environmental behaviors are inherently difficult to standardize and measure. Extant ESG rating indicators are predominantly designed with corporate legal entities as the unit of analysis, lacking compatibility with "non-corporatized" production entities. Existing study found that the average technical efficiency gap rate of ESG in the primary industry was merely 0.792, the lowest among all three sectors[1]. The European Union's "Farm to Fork" strategy and the United States' "Climate-Smart Agriculture" initiative have already incorporated smallholders into carbon footprint verification and ecological subsidy frameworks. The critical challenge lies in constructing measurement tools and incentive mechanisms adapted to their institutional environment.

Recent studies have examined agricultural ESG and smallholder green production from multiple perspectives. Existing work has established the relevance of digital transformation, fiscal support, sustainable certification, and green technology adoption[2-7], but it remains predominantly enterprise-centered. More importantly, studies of smallholder participation often accumulate multiple determinants without clearly distinguishing how different forms of support are translated into behavioral willingness. This leaves two related gaps: the lack of a smallholder-centered ESG framework and insufficient explanation of the cognitive mechanisms through which external support affects participation willingness.

To address these gaps, this study asks two questions: (1) How do key forms of external support and prior green behavior affect smallholders' willingness to participate in agricultural ESG practices? (2) Through what cognitive mechanisms do technical guidance and government subsidy translate into participation willingness? The study contributes in two ways. Theoretically, it extends agricultural ESG research from corporate organizations to smallholder agency and develops a mechanism framework. Practically, it provides evidence for designing technical services and subsidy instruments that enhance value recognition and alleviate perceived participation risks.

2. Literature Review

In recent years, scholarly attention has increasingly extended the ESG perspective to the agricultural supply chain domain. Existing research found that the environmental dimension exerts a significant positive effect on the scale expansion of agricultural enterprises[6]. Alareeni et al. identified differentiated impacts of ESG disclosure on firms' financial indicators[7]. Other literature demonstrated that ESG performance can significantly enhance corporate resilience through green innovation and reputation mechanisms[2]. Zhao et al. proposed that digital technologies can empower agricultural enterprises' ESG practices through three pathways: information transparency, process standardization, and transaction credibility[8]. Wu et al. verified the positive moderating effects of fiscal subsidies and media attention[3]. Existing literature found that the primary industry exhibited the lowest average ESG efficiency technical gap rate among all three industries[1]. However, some scholars argue that existing indicators lack compatibility with non-corporatized operational entities[9].

In terms of measurement instruments, the Sustainability Assessment of Food and Agriculture Systems (SAFA) framework developed by the Food and Agriculture Organization of the United Nations provides a systematic indicator system[10], yet its design serves formal organizational assessment and has limited applicability to smallholders. Existing research revealed that the removal of subsidies would lead to substantial declines in farmers' demand for key inputs[5]. Some scholars highlighted the significant potential of ICT in enhancing smallholders' resource utilization efficiency and environmental governance capacity[4].

Research on green technology adoption further enriches the understanding of smallholder participation. Prior studies proposed that a benefit-sharing mechanism based on "contract farming-cooperatives-farmers" can enhance participation[11], while some scholars highlighted the persistent gap between formal ESG standards and incentives reaching farmers[12]. Taken together, the literature suggests that smallholders respond not only to the presence of external support but also to how they interpret its value and risk implications. Therefore, rather than treating education, management, subsidies, technology, experience, and risk as a flat list of parallel predictors, this study concentrates on exploring how external support promotes the participation of small-scale farmers in ESG.

3. Research Methodology

This study adopted a mixed research design of "qualitative exploration-scale development-empirical testing". This design consists of two stages: (1) The qualitative stage - conducting semi-structured interviews with 35 small farmers, 10 agricultural management personnel, and 6 provincial and county-level agricultural department heads, using grounded theory for three-level coding to construct a theoretical framework; (2) The quantitative stage - developing a scale based on the qualitative findings, conducting a large-scale questionnaire survey of 506 participants, and using statistical models to test the hypotheses.

3.1. Qualitative Exploration: Grounded Theory Coding

This study selected three representative agricultural regions from Hebei Province, Liaoning Province and Shaanxi Province. The interview subjects comprised 35 smallholder farmers, 10 agricultural enterprise managers and staff, and 6 provincial and county-level agricultural authority personnel, totaling 35 hours of interview duration and yielding approximately 250,000 characters of interview transcripts.

This study rigorously followed the procedural grounded theory paradigm of Strauss and Corbin to conduct three-level coding of the interview transcripts[13]. Through open coding, 77 initial concepts were extracted from the raw statements and further categorized into 12 sub-categories. Axial coding consolidated these categories around three substantive elements: external support, cognitive appraisal, and participation willingness. Selective coding further clarified two differentiated cognitive pathways. Technical guidance helps farmers understand the practical and economic value of ESG practices, thereby strengthening ESG value perception; government subsidy reduces expected cost exposure and uncertainty, thereby strengthening perceived risk alleviation. Past green behavior emerged as an experiential foundation that can reinforce willingness directly.

This qualitative phase served as the exploratory foundation for the subsequent quantitative survey, following the exploratory sequential design advocated by Creswell and Plano Clark[14].

3.2. Scale Development and Questionnaire Design

This study developed the "Smallholder ESG Participation Behavior and Willingness Scale" based on the core categories extracted from grounded theory coding, including technical guidance, government subsidy, past green behavior, ESG value perception, perceived risk alleviation, and participation willingness. Education, gender, age, and land scale were retained as background controls. The questionnaire included items measuring both participation willingness and actual green behavior.

Following questionnaire design, two rounds of small-scale pilot testing (n = 30) were conducted. The formal survey employed a hybrid online-offline model: online, electronic questionnaires were generated via the Wenjuanxing platform and forwarded by cooperative leaders to farmer WeChat groups; offline, investigators conducted face-to-face household interviews. A total of 550 questionnaires were collected, and after data cleaning, 506 valid questionnaires were retained, yielding an effective response rate of 92.0%. Data entry employed dual-entry, cross-verification procedures and was subjected to dual validation using SPSS and Python to ensure consistent variable coding and standardized formatting. Table 1 presents the sample characteristics.

Table 1. Basic Sample Characteristics and Green Production Willingness Rate Analysis

Characteristic Dimension	Category	Sample Size	Proportion	Green Production Willingness Rate
Gender	Male	250	49.4%	43.6%
	Female	256	50.6%	39.5%
Age	≤ 30 years	83	16.4%	60.2%
	31–50 years	276	54.5%	35.5%
	≥ 51 years	147	29.1%	42.2%
Education	Primary school or below	78	15.4%	29.5%
	Junior high school	95	18.8%	42.1%
	High school / Vocational	208	41.1%	38.9%
	College or above	125	24.7%	52.8%
Land scale (mu) Note: 30 mu (2 hectares)	< 10	170	33.6%	49.4%
	10–30	178	35.2%	41.0%
	31–50	113	22.3%	32.7%
	> 50	45	8.9%	35.6%

3.3. Research Hypotheses and Variable Settings

(1) Theoretical Framework and Research Hypotheses

The theoretical model is organized around "stimulus-appraisal-intention" logic. External support does not necessarily translate directly into participation willingness; its effect depends on how smallholders cognitively interpret the support they receive. Consistent with the cognitive-affective-behavioral perspective, this study distinguishes two appraisal mechanisms. Technical guidance primarily reduces knowledge barriers and clarifies the usefulness of ESG practices, thereby enhancing ESG value perception. Government subsidy primarily reduces expected cost exposure and uncertainty, thereby strengthening perceived risk alleviation. In addition, experiential learning theory

suggests that prior green production experience can directly reinforce subsequent participation willingness.

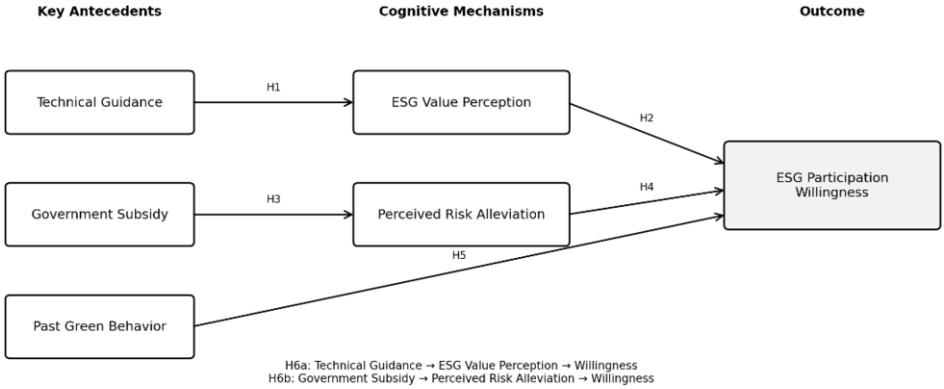


Figure 1. Theoretical Framework

Technical guidance converts abstract ESG standards into localized and actionable production knowledge. By improving farmers' understanding of the environmental, production, and potential economic benefits of ESG practices, effective guidance should strengthen their recognition of ESG value. Accordingly:

H1: Perceived technical guidance has a positive effect on smallholders' ESG value perception.

Value perception provides the cognitive basis for behavioral intention. When farmers believe that ESG practices generate meaningful environmental, production, or market value, they are more likely to form a willingness to participate. Accordingly:

H2: ESG value perception has a positive effect on smallholders' ESG participation willingness.

Government subsidy can offset additional input costs and signal institutional support, reducing farmers' concerns about financial loss and implementation uncertainty. Accordingly:

H3: Perceived government subsidy has a positive effect on smallholders' perceived risk alleviation.

Lower perceived to loss makes participation more acceptable. Thus, stronger perceived risk alleviation should increase willingness to engage in ESG. Accordingly:

H4: Perceived risk alleviation has a positive effect on smallholders' ESG participation willingness.

Past green behavior provides farmers with operational experience and feedback. Such experience reduces unfamiliarity and strengthens confidence. Accordingly:

H5: Past green behavior has a positive effect on smallholders' ESG participation willingness.

The preceding arguments imply two differentiated mediation mechanisms. Technical support operates primarily through value recognition, whereas economic support operates primarily through risk alleviation. Accordingly:

H6a: ESG value perception mediates the relationship between perceived technical guidance and ESG participation willingness.

H6b: Perceived risk alleviation mediates the relationship between perceived government subsidy and ESG participation willingness.

(2) Variable Settings

The dependent variable is smallholders' ESG participation willingness (Y).

The focal antecedent variables are perceived technical guidance (X1), perceived government subsidy (X2), and past green behavior (X3).

Mediating variables: ESG value perception (M1) and perceived risk alleviation (M2). M2 corresponds to the original reverse-coded risk-perception measure; higher values therefore indicate lower perceived risk and stronger risk alleviation.

Control variables: education level, gender, age, land scale, and unified management.

(3) Model Construction

This study employs a binary logistic regression model for empirical analysis. The model is specified as follows:

$$\log(\mathrm{P(Yi}=1)1-\mathrm{P(Yi}=1)) = \beta 0+\beta 1\mathrm{X1i}+\beta 2\mathrm{X2i}+\beta 3\mathrm{X3i}+\beta 4\mathrm{X4i}+\beta 5\mathrm{X5i}+\beta 6\mathrm{X6i}+\gamma 1\mathrm{Z1i}+\gamma 2\mathrm{Z2i}+\gamma 3\mathrm{Z3i}+\varepsilon i \log(1-\mathrm{P(Yi}=1)\mathrm{P(Yi}=1)) = \beta 0+\beta 1\mathrm{X1i}+\beta 2\mathrm{X2i}+\beta 3\mathrm{X3i}+\beta 4\mathrm{X4i}+\beta 5\mathrm{X5i}+\beta 6\mathrm{X6i}+\gamma 1\mathrm{Z1i}+\gamma 2\mathrm{Z2i}+\gamma 3\mathrm{Z3i}+\varepsilon i$$

P(Yi=1) denotes the probability that farmer i expresses willingness to participate in green production. The focal coefficients capture the effects of technical guidance, government subsidy, and past green behavior, while education, gender, age, and land scale are treated as controls. Unified management is included only in supplementary specifications. For mediation analysis, M1 and M2 are modeled as intervening cognitive mechanisms linking the corresponding external-support variable to Y.

4. Empirical Result

4.1. Direct Effects and Baseline Model

Given that participation willingness is measured on a five-point Likert scale, both binary logistic regression and ordered logistic regression were employed. The baseline model retains the original set of covariates for comparability, but interpretation focuses on the streamlined theoretical variables: technical guidance, government subsidy, past green behavior, and perceived risk alleviation, with education and demographic characteristics treated as controls and unified management treated as supplementary.

Table 2. Binary Logistic Regression Results (Baseline Model)

Variable	Coefficient (B)	SE	Wald z	p-value	OR [95% CI]
Tech. guidance	0.127	0.128	0.991	0.323	1.135 [0.883, 1.459]
Gov. subsidy	0.382	0.129	2.958	0.003**	1.465 [1.138, 1.887]
Past green beh.	0.705	0.137	5.162	0.000***	2.024 [1.548, 2.647]
Risk perception	0.691	0.131	5.282	0.000***	1.995 [1.543, 2.579]
Unified mgmt.	-0.015	0.118	-0.127	0.898	0.985 [0.782, 1.241]
Education	0.258	0.112	2.309	0.021*	1.294 [1.039, 1.611]
Land scale	-0.254	0.110	-2.306	0.021*	0.776 [0.625, 0.963]
Gender	0.361	0.206	1.754	0.081	1.435 [0.958, 2.149]
Age	-0.103	0.089	-1.153	0.249	0.902 [0.758, 1.074]
Intercept	-7.222	1.171	-6.169	0.000***	0.001

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. N = 506.

As shown in Table 2, past green behavior is the strongest focal direct predictor of participation willingness (B = 0.705, OR = 2.024, $p < 0.001$), supporting H5. Government subsidy also shows a significant positive direct association. The reverse-

coded risk measure is positively associated with willingness, consistent with H4 when interpreted as perceived risk alleviation. Technical guidance does not show a significant direct effect, which is consistent with the proposed mechanism model because its influence is expected to operate primarily through ESG value perception.

The absence of a direct effect for technical guidance does not imply that technical services are ineffective. Instead, it motivates the mechanism test: guidance may influence willingness only after farmers recognize the value of ESG practices. Unified management is not interpreted as a focal hypothesis because the streamlined framework treats it as a supplementary contextual factor rather than a core mechanism variable.

4.2. Mediation Effect Analysis

(1) OLS regression of the independent variable on the mediator

First, the effect of perceived technical guidance on ESG value perception (M1) was examined. Technical guidance has a positive and statistically significant coefficient (B = 0.276, $p < 0.001$), supporting H1.

Table 3. OLS Regression Results of Independent Variables on ESG Value Perception (M1)

Explanatory Variable	Coefficient (B)	SE	t-value	p-value	Goodness-of-fit / Sample Size
Perceived technical guidance	0.276	0.041	6.713	0.000***	F-statistic = 25.97***
Education level	0.096	0.043	2.221	0.027*	Adjusted $R^2 = 0.090$
Intercept	1.938	0.195	9.939	0.000***	N = 506

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. N = 506.

Second, the effect of perceived government subsidy on perceived risk alleviation (M2) was examined. Government subsidy has a positive and statistically significant coefficient (B = 0.323, $p < 0.001$), supporting H3.

Table 4. OLS Regression Results of Independent Variables on Risk Perception (M2)

Explanatory Variable	Coefficient (B)	SE	t-value	p-value	Goodness-of-fit / Sample Size
Perceived government subsidy	0.323	0.030	10.885	0.000***	Adjusted $R^2 = 0.187$
Past green behavior	−0.040	0.038	−1.050	0.294	F-statistic = 59.24***
Intercept	2.247	0.171	13.131	0.000***	N = 506

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. N = 506.

(2) Bootstrap Mediation Effect Test

Table 5 reports the Bootstrap test results (1,000 resamples) for the four mediation pathways.

Table 5. Mediation Effect Test Results (Bootstrap, 1,000 Resamples)

Mediation Pathway	Mediation Effect (ab)	95% Confidence Interval	p-value	Conclusion
X1 → M1 → Y	0.235	[0.140, 0.367]	0.000***	H6a supported
X2 → M2 → Y	0.121	[0.008, 0.257]	0.032*	H6b supported

Note: A confidence interval that does not contain zero indicates a significant mediation effect. * $p < 0.05$, *** $p < 0.001$.

First, ESG value perception significantly mediates the relationship between perceived technical guidance and participation willingness. The result indicates that technical guidance operates through value recognition. H2 and H6a are supported.

Second, perceived risk alleviation significantly mediates the relationship between government subsidy and participation willingness. The research result indicates a partial mediation mechanism: subsidy both improves willingness directly and indirectly by alleviating perceived risk. H4 and H6b are supported.

4.3. Robustness Tests

To verify the robustness of the baseline regression results, this study employed an ordered logistic regression model. The results of the ordered logistic regression, reported in Table 6, are consistent with the binary logistic regression in terms of the direction and significance of the explanatory variables, indicating that the baseline model's findings are robust to alternative model specifications.

Table 6. Ordered Logistic Regression Results (Robustness Test)

Variable	Coefficient (B)	SE	Wald z	p-value	Consistency
Tech. guidance	0.127	0.128	0.991	0.323	Consistent
Gov. subsidy	0.382	0.129	2.958	0.003**	Consistent
Past green beh.	0.705	0.137	5.162	0.000***	Consistent
Risk perception	0.691	0.131	5.282	0.000***	Consistent
Unified mgmt.	-0.015	0.118	-0.127	0.898	Consistent
Education	0.258	0.112	2.309	0.021*	Consistent
Land scale	-0.254	0.110	-2.306	0.021*	Consistent
Gender	0.361	0.206	1.754	0.081	Consistent
Age	-0.103	0.089	-1.153	0.249	Consistent
Intercept	-7.222	1.171	-6.169	0.000***	Consistent

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. N = 506.

5. Conclusions and Discussion

5.1. Conclusions

This study combines qualitative data from 51 interviewees with survey data from 506 smallholder farmers to explain how external support and prior behavioral experience shape agricultural ESG participation willingness. Rather than treating multiple determinants as parallel predictors, the revised framework distinguishes two cognitive mechanisms - ESG value perception and perceived risk alleviation - and one direct experiential pathway.

First, descriptive statistics reveal that in terms of reducing chemical pesticide use, only 20.2% of farmers fully complied with the requirements, and in terms of increasing organic fertilizer use, only 18.2% fully complied, while the proportions of those who did not comply at all were 14.2% and 12.3%, respectively. Although smallholders possess participation willingness, their actual implementation rate remains at a low level.

Second, past green behavior is the strongest direct predictor of participation willingness, indicating that firsthand experience provides an important experiential foundation for continued ESG engagement. Government subsidy also has a significant positive direct association with willingness.

Third, external support operates through differentiated cognitive mechanisms. Technical guidance influences willingness primarily by enhancing ESG value perception, whereas government subsidy promotes willingness partly by strengthening perceived risk alleviation. This distinction between a value-recognition pathway and a risk-

alleviation pathway provides the central mechanism-based explanation of smallholder ESG participation.

5.2. Contribution

This study makes three theoretical contributions. First, it extends agricultural ESG research from corporate organizations to smallholder agency, addressing a critical governance link at the base of the agricultural supply chain[15]. Second, it develops and validates a smallholder-oriented measurement framework based on grounded theory and first-hand survey data. Third, and most importantly, it moves beyond a flat list of participation determinants by identifying differentiated cognitive pathways: technical guidance works through ESG value recognition, while government subsidy works partly through perceived risk alleviation. The independent effect of past green behavior further shows that experiential reinforcement constitutes a distinct behavioral foundation. Together, these findings provide a more parsimonious explanation of how external support is translated into participation willingness[16].

This study also makes practical contributions. First, implement differentiated subsidies for organic fertilizers and biopesticides, with a priority preference for small-scale farmers. Additionally, establish a first-year risk compensation fund in conjunction with agricultural insurance to reduce farmers' trial-and-error costs. Second, develop practical operational atlases for ESG technologies tailored to staple crops at the county level, deploy local technical personnel through cooperatives, and conduct regular field visits. Technical services should integrate operational guidance with value clarification, enabling farmers to first believe that the practices are worthwhile and then learn how to implement them. Third, establish county-level agricultural product traceability and premium pricing platforms, create origin traceability labels for ESG-certified agricultural products, and link them with e-commerce and supermarket channels. Simultaneously, encourage leading agricultural enterprises and cooperatives to sign medium-to-long-term guaranteed-price procurement contracts to stabilize income expectations, thereby translating green behaviors into tangible economic returns.

5.3. Limitations

The survey sample is drawn exclusively from Hebei and Liaoning provinces, which may limit the generalizability of the findings to other agricultural regions with distinct production systems and policy environments. Our results are most applicable to the North China Plain and Northeast China; application to southern cash-crop areas or western ecologically fragile zones requires caution. The cross-sectional design precludes causal inference regarding the temporal dynamics of the willingness–behavior relationship. Moreover, our policy recommendations, though grounded in significant variables, have not been tested through randomized interventions or simulation; they remain correlational inferences rather than causal evidence.

Future research should extend the geographic scope, adopt longitudinal or panel designs to track the evolution of smallholder ESG participation, and explore the moderating roles of digital technology adoption and cooperative organizational forms in the conversion from willingness to actual behavior. Experimental or quasi-experimental studies, e.g., randomized subsidy trials, training program evaluations, are particularly encouraged to provide stronger causal evidence for intervention effectiveness.

References

[1] Wang H Q. Static and dynamic evolution of ESG efficiency across China's three major industries: An analysis based on meta-frontier theory and Malmquist index. Management and Technology of SMEs, 2024, (13).
[2] Jin S R, Qi X H, Xiong R Y. ESG performance and agricultural enterprise resilience: Impact effects and mechanisms. Reform, 2025, (3), 124–141.
[3] Wu J Y, Zhao C X, Zhang J. ESG performance and agricultural leading enterprise performance: A study based on green development and media spotlight effects. Commercial Accounting, 2025, (13).
[4] Bertacchini F, Cimino A, Fusto C, et al. ICTs and smallholders: A systematic review of case studies and their impact on ESG outcomes. Procedia Computer Science, 2025, 253, 3276–3287.
[5] Lal P, Chandel B S, Tiwari R K, et al. Effects of agricultural subsidies on farm household decisions: A separable household model approach. Frontiers in Sustainable Food Systems, 2023, 7.
[6] Cristea M, Noja G G, Thalassinos E, et al. Environmental, social and governance credentials of agricultural companies—The interplay with company size. Resources, 2020, 11(3), 30.
[7] Alareeni B A, Hamdan A. ESG impact on performance of US S&P 500-listed firms. Corporate Governance: The International Journal of Business in Society, 2020, 20(7), 1409–1428.
[8] Zhao X L, Hou M J. Theoretical logic and practical pathways of digital transformation empowering corporate ESG practice: A case study of agricultural enterprises. China Business Review, 2025, 34(4), 155.
[9] Zhang Y T. Construction and application of ESG performance evaluation system in the aquaculture industry: A case study of Wens Group. Heilongjiang Bayi Agricultural University, 2025.
[10] FAO. Sustainability assessment of food and agriculture systems (SAFA): Guidelines (Version 3.0). Food and Agriculture Organization of the United Nations. 2014.
[11] Qian K. Exploring sustainable development models of urban pastoralism from an ESG perspective: A case study of the Yuetang East area in Ningbo. Building Energy Efficiency (Chinese and English), 2025, 53(8), 158–162.
[12] Liu R X. Dilemmas and countermeasures of ESG practice in agricultural economic sustainability. Shanghai Business, 2025, (1), 235–237.
[13] Strauss A, Corbin J. Basics of qualitative research: Grounded theory procedures and techniques. Sage Publications. 1990.
[14] Creswell J W, Clark V L P. Designing and conducting mixed methods research. Sage, 2007.
[15] Nuraisyah A, Wulandari E, Indrawan D, et al. A Use-Case Analytical Review of Supply Chain Integration and Sustainable Financing in Agricultural Innovation. Supply Chain Analytics, 2026, 100220.
[16] Cao H, Li F, Zhao K, et al. From value perception to behavioural intention: Study of Chinese smallholders' pro-environmental agricultural practices. Journal of Environmental Management, 2022, 315, 115179.

Section 5

Interdisciplinary Applications of Digitalization and Management Innovation

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261136

Research on the Influencing Factors of the Number of Collective Labor Disputes

Haichao Lu[1], Binbin Hu and Xiuyun Gong
Beijing Wuzi University, China

Abstract. This study explores the macro-environmental factors influencing changes in the number of collective labor disputes in China, aiming to provide empirical evidence for the prevention and resolution of such disputes. Based on the Anderson Model, a five-dimensional analytical framework encompassing the legal, economic, political, socio-cultural, and technological dimensions was established. Using panel data from 29 provinces in China spanning 2004 to 2023, an empirical analysis was conducted via the two-way fixed effects model. The results show that: the amendment to the Labor Contract Law significantly inhibited collective labor disputes; regional gross domestic product (GDP), the number of participants in unemployment insurance, the success rate of labor dispute mediation, and the average years of schooling were significantly positively correlated with the number of disputes; the average wage of employed personnel and full-time equivalent of R&D had no significant impact at the national level, while technological progress in the central region exerted a significantly positive impact on disputes.

Keywords. Collective labor disputes; Anderson model; two-way fixed effects; labor dispute

1.Introduction

1.1. Research Background and Significance

The harmony and stability of labor relations are a core cornerstone of social development. Amid global economic restructuring, interest pattern adjustments, and workers' growing rights awareness, labor relations contradictions have become increasingly prominent, with collective labor disputes—an important and special form of labor dispute—deserving special attention.

Compared with individual labor disputes, collective labor disputes involve group conflicts between multiple workers and employers, featuring greater complexity, wider social impact, and potential destructiveness. They often reflect deep-seated systemic and structural issues in enterprises (e.g., management, compensation, working conditions), and may lead to production disruptions, enterprise losses, social instability, and the consumption of administrative and judicial resources.

In recent years, the quantity and influence of collective labor dispute cases have been on the rise, exposing enterprise management deficiencies and the gap between workers' rights protection demands (for better treatment and working conditions) and

[1] Corresponding Author, Haichao Lu, Beijing Wuzi University, China; Email: 1372336703@qq.com

enterprise practices. Systematically exploring the key factors affecting the number of such disputes thus holds significant theoretical and practical value: it provides a scientific basis for governments to formulate targeted labor policies and laws, guides enterprises in optimizing internal management and building inclusive labor relations, and ultimately promotes the sustainable and healthy development of the economy and society.

1.2 Literature Review and Research Review

Existing studies have explored the influencing factors of collective labor disputes from multiple dimensions−ranging from macro-level economic environments and labor market institutions to micro-level organizational internal factors−confirming that collective labor disputes are complex social phenomena driven by the joint action of multiple factors. Below is a systematic collation of foreign and domestic research results, followed by a review of the strengths and gaps of existing studies, and a clarification of the contributions of this research.

1.2.1 Foreign Research Review

Foreign scholars focus on the interaction between labor relations systems, institutional environments, and organizational practices, and their research can be categorized into four core directions:

(1) Influence of Labor Relations Nature and Employment Models. Fahn [1] analyzed the role of "reciprocity" in dynamic employment relationships in Reciprocity in Dynamic Employment Relationships, pointing out that imbalances in reciprocity (e.g., in compensation, welfare, or working conditions) can accumulate employee dissatisfaction and increase the risk of collective disputes. Rogovsky et al. [1] studied labor relations in the taxi industry in Employment Relationships in Taxi Organizations: Linking Agency to Firm Capabilities and found that specific employment models (such as the agency model) lead to information asymmetry and interest distribution conflicts between labor and management, which easily trigger collective disputes.

(2) Influence of Labor Market Institutions and Macroeconomic Environment. Buchele & Christiansen [2] pointed out in Employment and Productivity Growth in Europe and North America: The Impact of Labor Market Institutions that labor market institutions (e.g., collective bargaining mechanisms, employment security systems) have a dual impact on employment and productivity. Specifically, the lack of effective collective bargaining or insufficient employment security tends to make employees express dissatisfaction through collective disputes, while sound institutions can guide conflicts toward non-confrontational resolution. Vieriu & Vieriu [3] further emphasized in General Aspects of the Evolution of Regulations on Collective Labor Disputes in Romania that the evolution of labor laws directly affects the number of collective disputes: clear, fair, and enforceable laws can regulate the behavior of labor and management, and reduce conflicts caused by ambiguous rules or ineffective enforcement.

(3) Internal Organizational Factors and Communication Mechanisms. Hui et al. [4] examined Chinese workers' organizational attachment in Employment Relationships in China: Do Workers Relate to the Organization or to Individuals? and found that low employee identification with the organization or lack of trust in management can easily escalate to collective problems, thereby triggering disputes. Guț Van [5] verified the

key role of communication in The Role of Effective Communication in Preventing Labor Conflicts within Local Public Administration Institutions: the absence of transparent and open communication channels leads to the accumulation of dissatisfaction, while unobstructed communication can resolve problems in a timely manner and reduce collective disputes.

(4) Conflict Resolution Mechanisms and Alternative Dispute Resolution (ADR).Volosevici [6] and Simone[7] argued in Alternative Dispute Resolution and Individual Labor Disputes: A Different Approach to Organizational Dispute Management and Labour Conflicts in the Digital Age-a Comparative Perspectiveby Donatelladella Porta, Riccardo Emiliochesta and Lorenzocini that ADR mechanisms (e.g., collective negotiation, mediation, arbitration) are effective ways to prevent the escalation of collective disputes. If ADR mechanisms are inefficient or lack credibility, collective disputes are more likely to evolve into confrontational forms such as strikes, increasing their quantity and negative impact.

1.2.2 Domestic Research Review

Domestic research focuses on the characteristics of collective labor disputes in China's economic transition period and the influence of institutional and economic factors, with research contents concentrated in four areas:

(1) General Trends and Characteristics of the Number of Collective Labor Disputes. Most studies confirm an overall upward trend in collective labor disputes. Wang Zhiwei et al. [8] analyzed data from 1991 to 2016 and found that the number of collective disputes continued to rise, with enterprises as the main responsible party, and infringements on labor remuneration and social insurance rights as the primary causes. Kong Lingming [9] categorized the causes of 3,110 collective disputes in City D into seven types (e.g., wage arrears, demands for improved working conditions) and summarized five worker action models (e.g., petitions, strikes). Some studies also note volatility: Gao Hongyan [10] pointed out that labor relations have become more confrontational during the economic transition; Zhan Yubo & Chen Qiang [11] used inter-provincial panel data (1998−2012) to find that foreign trade (especially processing trade) significantly increases disputes, while trade unions have a limited effect on reducing disputes.

(2) Economic Development Factors. Zhang Shaoping [12] proposed a correlation between macroeconomic indicators and labor disputes. Wang Yang [13] found that economic restructuring, system transition, and scale expansion intensify labor-management tensions, while economic development level and growth rate alleviate tensions to a certain extent, but the impact of unemployment rate and education level is regionally heterogeneous. Cheng Xiao & Li Jing [14] further revealed a chain reaction: the number of disputes is positively correlated with per capita capital stock and negatively correlated with labor income share; disputes promote enterprises to replace labor with capital, and capital deepening plays a mediating role. Hu Lei [15]found an "inverted U-shaped" relationship between economic development level and dispute incidence (1991−2011 data): disputes increase in the early stage of development and stabilize or decline after reaching a threshold.

(3) Legal System Factors. Yu Qin et al.[16]analyzed the 40-year evolution of labor legislation and dispute handling, and found that the improvement of the legal system affects the efficiency of dispute resolution and workers' willingness to protect rights. Chen Tianxue & Zhao Man [17] emphasized that labor law adjustments directly

affect the number of disputes and are key to predicting dispute quantities. Regarding resolution mechanisms, Zhuang Wenjia [18] used panel data (1999−2011) to find that strong mediation capabilities of labor arbitration institutions reduce the scale of collective dispute participants but not their frequency; Luo Yan & Lin Qiulan [19] confirmed that democratic participation (e.g., employee rationalization proposals) and mediation committees are negatively correlated with the number of disputes. However, Chang Cheng [20] questioned the applicability of mediation mechanisms through 106 cases in Beijing's Xicheng District, arguing that weak trade unions and softened legal standards may lead to secondary infringements on workers' rights.

(4) Labor Market and Social Management Factors. Wang Caiping pointed out that the number of enterprise group disputes increases year by year, and improving collective negotiation mechanisms and trade union systems is crucial for prevention. Kong Lingming [21] found that the 2008 international financial crisis triggered a surge in wage arrears disputes, while the criminalization of wage arrears (since 2013) gradually reduced such cases. Regional differences are also notable: Chen Chong & He Mengxing [22] analyzed Northwest China and found that economic scale, labor market conditions, and marketization level have heterogeneous impacts on disputes.

1.2.3 Review of Existing Research and Contributions of This Study

(1) Strengths of Existing Research

1. Multi-dimensional exploration: Studies cover macro factors (economic environment, legal system) and micro factors (organizational communication, trade union role), forming a relatively comprehensive research scope.

2. Quantitative method application: Domestic studies widely use panel data analysis and time series analysis to test the relationship between disputes and economic development, capital deepening, etc., improving research scientificity.

3. In-depth analysis of specific factors: Studies on the inhibitory effect of wage arrears criminalization and the role of ADR mechanisms provide targeted empirical evidence for policy formulation.

4. Focus on new challenges: Cutting-edge research pays attention to the impact of new business formats (e.g., platform economy) on labor relations, reflecting the contemporaneity of the research.

(2) Gaps in Existing Research

1. Lack of a systematic integrated framework: Most studies focus on one or two dimensions (e.g., economic or legal factors) and fail to integrate legal, economic, political, socio-cultural, and technological factors into a unified analytical framework, leading to a fragmented understanding of factor interactions.

2. Insufficient control of confounding factors: When using provincial panel data, few studies adopt the two-way fixed effects model (which controls for time-invariant regional characteristics and region-invariant time shocks), resulting in potential omitted variable bias and reduced conclusion reliability.

(3) Contributions of This Study

1. Theoretical contribution: By introducing the Anderson Model, this study constructs a "five-in-one" comprehensive analytical framework covering the legal, economic, political, socio-cultural, and technological environments, overcoming the single perspective limitation of previous studies and providing an integrated theoretical perspective for understanding the driving mechanism of collective disputes.

2. Methodological contribution: This study uses the two-way fixed effects model to conduct regression analysis on inter-provincial panel data (2004−2023, 29 provinces). By controlling for regional and time fixed effects, it accurately identifies the causal impact of core variables, reduces omitted variable bias, and enhances the robustness of conclusions.

2.Theoretical Analysis

2.1. Anderson Model

Dunlop's theoretical framework has undergone several evolutions, among which the Anderson Model recognizes that the "inputs" of the labor relations system include various other factors beyond market, technology, and power. This model views the labor relations system as a dynamic system and reveals a causal relationship within it: the output at any level (enterprise, industry, sector, society) or any period may serve as the input for other levels or periods.

The Anderson Model also clearly distinguishes between procedural rules (conversion process) and substantive rules (output). It argues that the network of substantive rules includes not only wages, welfare, and working conditions but also other outputs of the conversion mechanism, such as laws, productivity, labor-management conflicts, managerial authority, resignations, absenteeism, and employees' work attitudes.

Drawing on the analytical logic of the Anderson Model, this study regards the number of collective labor disputes as the "output" of the labor relations system and constructs a "five-in-one" comprehensive analytical framework to thoroughly explore the multidimensional "inputs" and "conversion mechanisms" that influence this output. The framework treats the labor relations system as an open, dynamic system, whose output is comprehensively affected by the following environmental factors:

(1) Economic Environment. Macroeconomic variables such as macroeconomic fluctuations (e.g., economic growth rate, unemployment rate, inflation), industrial structure adjustment, and capital deepening directly affect enterprise operating conditions, employees' job stability, and income levels—thereby triggering or exacerbating labor-management conflicts. Multiple domestic studies have pointed out that close correlations exist between the level of economic development, economic structure adjustment, and the number of labor disputes.

(2) Legal Environment. The sophistication of labor laws and regulations, the intensity of their enforcement, and the effectiveness of dispute resolution mechanisms directly influence workers' cost-benefit of safeguarding rights and the efficiency of dispute resolution. For example, the evolution of laws and regulations affects the resolution methods and quantity of collective labor disputes. The "mediation first" principle and the mediation capabilities of labor arbitration institutions significantly reduce the scale of participants in collective labor disputes; meanwhile, sound laws and regulations can effectively regulate the behaviors of labor and management, reducing collective conflicts caused by ambiguous rules or ineffective enforcement.

(3) Social and Cultural Environment. Factors such as social concepts of equity, enhanced awareness of workers' rights, changes in social management models, and historical and cultural backgrounds all shape workers' behavioral patterns and

tendencies toward collective action. For instance, low employee identification with the organization or lack of trust in management can easily lead to problems at the collective level, thereby triggering collective disputes. At the same time, contextual factors such as ethnicity and class struggle may also affect labor conflicts.

(4) Political Environment. The government's role positioning in labor relations, the orientation of labor policy formulation, and the support or restrictions on workers' organizations (e.g., trade unions) profoundly influence the balance of power in labor relations and the occurrence of disputes. For example, effective collective bargaining mechanisms can provide channels for dispute resolution, thereby reducing the number of dispute outbreaks.

(5) Technological Environment. Emerging technologies such as digitalization and the platform economy have brought structural changes to labor organization forms, employment models, and employees' job stability. These changes may spawn new types of collective labor disputes and pose challenges to traditional collective bargaining and dispute resolution mechanisms. Although digital technology has created new forms of employment, it may also exacerbate the fragmentation and instability of the labor force.

2.2. Analytical Framework

In the analytical framework of this study, the five aforementioned environmental factors serve as the independent variables influencing the number of collective labor disputes. The conversion process within the labor relations system—including collective negotiation, the role of trade unions, internal enterprise communication mechanisms, and the operational efficiency of alternative dispute resolution (ADR) mechanisms—determines how these inputs are transformed into the final output, i.e., the number of collective labor disputes.

The Anderson Model's emphasis on the distinction between procedural rules (conversion process) and substantive rules (output), as well as the causal relationships it reveals, provides theoretical support for this study to understand how inputs affect outputs through conversion mechanisms.

By constructing this "five-in-one" analytical framework and using the two-way fixed effects model for empirical testing, this study adopts a quantitative research approach to comprehensively and systematically identify and evaluate the impact of various environmental factors on the number of collective labor disputes. This aims to provide more precise theoretical guidance and policy recommendations for the effective prevention and resolution of collective labor disputes.

3. Empirical Analysis

3.1. Indicator Selection and Model Construction.

Drawing on Anderson's system theory of labor relations, this study classifies the factors influencing the number of collective labor disputes into five environmental systems: legal, economic, political, socio-cultural, and technological. Based on this framework and considering data availability, relevant indicators are selected to construct a panel data model for empirical testing.

3.1.1 Selection of Dependent Variable Indicator

The core focus of this study is to explore the factors affecting the number of collective labor disputes. Therefore, the dependent variable in this paper is the number of collective labor dispute cases.

This indicator directly reflects the frequency of group conflicts between labor and management in the labor relations system that have intensified and entered formal mediation or arbitration procedures. Compared with individual labor disputes, collective labor disputes involve more people and exert greater social impact, serving as a key "output" indicator for measuring the tension of labor relations and systemic risks in a region.

3.1.2 Selection of Independent Variable Indicators

Considering the scientificity and availability of data, the following indicators are selected:

(1) Legal Environment. Legal Implementation (law): This study uses a dummy variable for measurement. On July 1, 2013, the revised Labor Contract Law was formally implemented, introducing stricter provisions on employment forms such as labor dispatch. This represented a significant change in the legal environment of labor relations. Therefore, this paper assigns a value of "1" to the years 2013 and beyond, and "0" to the years before 2013, to capture the potential impact of this legal policy change on collective labor disputes.

(2) Economic Environment. Following the approaches of Wu Yuming & Zhang Yan and Zhou Cheng et al. , two indicators—economic scale and average wage—are selected to measure and represent the economic environment. Economic Scale: Measured by the Gross Regional Product (GRP) of each province. Economic scale not only reflects a region's economic strength but is also closely related to enterprise operating conditions, employment capacity, and salary levels, serving as the macro background that determines the interest distribution and bargaining power of both labor and management. Average Wage: Measured by the average wage of urban employed persons in each province. Labor remuneration is one of the primary triggers of labor disputes. Changes in the average wage level may affect workers' satisfaction and expectations, thereby influencing the frequency of collective labor disputes.

(3) Political Environment. Success Rate of Labor Dispute Mediation (mediate): Measured by the ratio of "cases successfully mediated by labor dispute mediation committees" to "total accepted cases". This is a key indicator for measuring the effectiveness of dispute resolution mechanisms in the "conversion process" of labor relations. Social Security Level (unemploy): Measured by the number of participants in unemployment insurance in each province. Social security is an important component of labor relations and a high-frequency area for labor disputes. As a key part of the social safety net, the coverage of unemployment insurance (reflected by the number of participants) indicates the basic security level for workers.

(4) Socio-Cultural Environment. Referring to the calculation method of the "Cultural Education and Social Security" dimension in the National Bureau of Statistics' Regional Development and Livelihood Index, the average years of schooling is selected to measure the socio-cultural environment.

Average Years of Schooling (edu): Measured by the average years of schooling of the labor force in each region. Workers' educational level is an important reflection of their human capital and awareness of social rights. On the one hand, an increase in

educational level may enhance workers' awareness and ability to safeguard their rights using legal means, thereby increasing the number of disputes; on the other hand, it may also improve their ability to communicate and negotiate with management to resolve issues, thereby reducing unnecessary conflicts.

(5) Technological Environment. Following Yu Wei et al. and the calculation method for the "Technological Innovation" section in the National Bureau of Statistics' Regional Development and Livelihood Index, R&D full-time equivalent is selected to measure the technological environment. Investment in Scientific and Technological Human Capital (rds): Measured by R&D (Research and Experimental Development) full-time equivalent data. The intensity of R&D investment reflects the level of technological progress and industrial upgrading in a region. Technological changes profoundly affect employment models, work organization forms, and the balance of power between labor and management, which may give rise to new forms of labor relations and potential conflict points—making it an unavoidable factor influencing modern labor disputes.

The specific variable names and explanations are detailed in the Table 1:

Table 1. Names and Definitions of Variables

Variable Type		Variable Name	Variable Code
Dependent Variable		Number of collective labor dispute cases(case)	case1
Independent Variable	Economic Environment	Gross Regional Product (100 million yuan)	gdp
		Average wage of employed persons (yuan)	wage
	Political Environment	Success rate of labor dispute mediation	mediate
		Number of unemployment insurance participants (10,000 persons)	unemploy
	Socio-Cultural Environment	Average years of schooling (year)	edu
	Technological Environment	R&D full-time equivalent (person-year)	rds
Dummy Variable	Legal Environment	Whether the Labor Contract Law is implemented	law

3.1.3 Data Sources

This study uses provincial-level panel data from 29 provinces in China from 2004 to 2023 for empirical testing. The Tibet Autonomous Region and the Xinjiang Production and Construction Corps are excluded due to significant missing data for the relevant indicators involved. All data used in this paper are derived from the panel data of the China Statistical Yearbook, China Labor Statistical Yearbook, and China Science and Technology Statistical Yearbook for 2004–2023, with relatively complete data coverage.

3.1.4 Model Construction

To reduce data volatility and eliminate partial heteroscedasticity, this paper takes the natural logarithm or conducts standardization on the dependent variable and independent variables with large fluctuations. The natural logarithm of the number of

collective labor dispute cases (lncase1) is used as the dependent variable, while the natural logarithm of Gross Regional Product (lngdp), the natural logarithm of the average wage of employed persons (lnwage), the success rate of labor dispute mediation (mediate), the standardized number of unemployment insurance participants (unemploy_std), the natural logarithm of average years of schooling (lnedu), the standardized R&D full-time equivalent (rds_std), and the dummy variable for the implementation of the Labor Contract Law (law) serve as independent variables. A two-way fixed effects regression model is established as follows:

$$\text{Incasel}_{i,t} = \alpha_{i,t} + \beta_i X_{i,t} + \gamma_i + \delta_t + \varepsilon_{i,t} \quad (1)$$

In Equation (1): $\alpha_{i,t}$ denotes the intercept term; β_i represents the coefficient of each independent variable, $X_{i,t}$ is the matrix of independent variables; γ_i and δ_t respectively represent the regional-specific effects and year-specific effects in the model; $\varepsilon_{i,t}$ is the random error term.

3.2. Empirical Results and Relevant Tests

3.2.1 Baseline Regression

Tests show that the Variance Inflation Factor (VIF) values of all variables in the model are less than 10, with an average VIF of 4.79, indicating no significant multicollinearity in the model and the results are shown in Table 2.

Table 2. VIF Test Results

	lncase1
law	-0.371**
	(-3.07)
lngdp	0.450***
	(8.13)
unemploy_std	0.607***
	(5.87)
lnwage	-1.225***
	(-9.50)
rds_std	0.0315
	(0.31)
mediate	-0.0450
	(-0.25)
lnedu	2.040***
	(5.10)
_cons	9.554***
	(9.04)
N	580

Subsequently, a cross-sectional dependence test and a Hausman test are conducted to determine whether to apply a fixed effects regression model, random effects regression model, or pooled OLS regression model to the data. The Friedman test yields a result of Pr = 1.0000, failing to reject the null hypothesis, indicating no cross-sectional dependence in the model. In the fixed effects regression, the F-statistic

is 39.8426, with the corresponding p-value significant at the 1% level, suggesting that the Fixed Effects (FE) model is significantly superior to the OLS regression. The Lagrange Multiplier (LM) test yields a chi-bar result of 672.01, with the p-value significant at the 1% level, indicating that the Random Effects (RE) model is significantly superior to the OLS regression. The Hausman test yields a result of chi2(7) = 87.36, with the corresponding p-value less than the 1% significance level. Therefore, the null hypothesis is rejected, indicating that the fixed effects model should be adopted to control individual heterogeneity related to the independent variables. The specific regression results are shown in Table 3.

Table 3. Comparison of Baseline Regression Results

	(1) OLS	(2) FE
law	-0.3712**	0.1900
	(0.1211)	(0.1031)
lngdp	0.4502***	0.9704***
	(0.0554)	(0.2222)
unemploy_std	0.6074***	0.4977***
	(0.1034)	(0.1360)
lnwage	-1.2247***	-1.6875***
	(0.1289)	(0.2439)
rds_std	0.0315	-0.1023
	(0.1018)	(0.1066)
mediate	-0.0450	0.2106
	(0.1812)	(0.1604)
lnedu	2.0396***	-1.9503*
	(0.4003)	(0.8444)
_cons	9.5544***	18.1735***
	(1.0568)	(1.4338)
N	580	580
F	118.8196	39.8426

Standard errors in parentheses

* p < 0.05, ** p < 0.01, *** p < 0.001

After confirming the use of the fixed effects model, time dummy variables are introduced to control for time effects in the model, thereby capturing effects such as technological progress and the details are shown in Table 4.

Table 4. Comparison of Regression Results with Time Effects Controlled

	(1) lncase1	(2) lncase1
law	0.1900	-3.2537**
	(0.1031)	(1.1332)
lngdp	0.9704***	1.0929***
	(0.2222)	(0.2275)
unemploy_std	0.4977***	0.5616***
	(0.1360)	(0.1291)
lnwage	-1.6875***	-0.5846
	(0.2439)	(0.6055)

rds_std	-0.1023	-0.1090
	(0.1066)	(0.1015)
mediate	0.2106	0.4595**
	(0.1604)	(0.1677)
lnedu	-1.9503*	2.2349*
	(0.8444)	(0.9860)
_cons	18.1735***	-2.4759
	(1.4338)	(5.5140)
Region	√	√
Time		√
N	580	580
R^2	0.339	0.452
adj. R^2	0.296	0.396
F	39.8426	17.3250
p	0.0000	0.0000

Standard errors in parentheses

* p < 0.05, ** p < 0.01, *** p < 0.001

Tests show that the F-statistic of all year dummy variables is 6, significant at the 1% level. Therefore, time fixed effects should be included in the model, and a two-way fixed effects model controlling for time effects is used for regression.

The regression results indicate that, after controlling for time and regional effects, the average wage level and the full-time equivalent of R&D personnel (rds_std) exert no statistically significant impact on the number of collective labor dispute cases (the explained variable), whereas all other explanatory variables exhibit a significant correlation with the explained variable. The specific empirical findings are as follows:

Regarding the legal environment, the regression coefficient of the dummy variable (law) representing the revision of the Labor Contract Law is -3.2537, which is significant at the 1% statistical level (p=0.004). This suggests that the implementation of this revision has significantly reduced the number of collective labor dispute cases by approximately 3%.

In terms of the economic environment, the regression coefficient of lngdp is 1.0929, highly significant at the 0.1% statistical level (p=0.000). This reveals a significant positive association between the expansion of regional economic scale and the number of collective labor dispute cases—specifically, a 1% increase in gross regional product leads to an average 1.09% rise in the number of such cases. In contrast, the regression coefficient of lnwage is -0.5846 (p=0.335), indicating no statistically significant impact. For the political environment, the regression coefficient of unemploy_std is 0.5616, highly significant at the 0.1% statistical level (p=0.000). A 0.01-standard-deviation increase in this variable results in an average 0.56% increase in the number of collective labor dispute cases. Additionally, the regression coefficient of mediate is 0.4595, significant at the 1% statistical level (p=0.006), meaning that a 1% increase in the mediation success rate leads to an average 0.46% rise in the number of such cases. In the sociocultural context, the regression coefficient of lnedu is 2.2349, significant at the 5% statistical level (p=0.024). This demonstrates a significant positive correlation between laborers' average years of schooling and the number of collective labor dispute cases: a 1% increase in average years of schooling causes an average 2.23% increase in the number of such cases. With respect to the technological

environment, the regression coefficient of rds_std is -0.1090 (p=0.283). This indicates that changes in the technological environment—measured by the investment intensity in science and technology human capital—have no significant impact on the number of collective labor dispute cases under the sample and model specification of this study.

3.2.2 Robustness Test

To verify the reliability and robustness of the model results, this paper conducts robustness tests from different perspectives, including replacing the dependent variable, excluding municipalities directly under the Central Government, and excluding special years. Through these tests, we aim to confirm whether the direction, significance, and impact magnitude of the coefficients of core variables remain consistent under different conditions, thereby ensuring the robustness and generalizability of the research conclusions. The specific robustness test results are shown in Table 5:

Table 5. Robustness Test Results

	(1) Original Regression	(2) Replaced Dependent Variable	(3) Excluded Special Year	(4) Excluded Municipalities
law	-3.2537**	-2.7903*	-3.2712**	-4.1509***
	(1.1332)	(1.0823)	(1.1600)	(1.1700)
lngdp	1.0929***	0.9379***	1.1675***	1.2050***
	(0.2275)	(0.2172)	(0.2385)	(0.2415)
unemploy_std	0.5616***	0.4392***	0.5651***	0.5837***
	(0.1291)	(0.1233)	(0.1322)	(0.1393)
lnwage	-0.5846	-0.6816	-0.6831	-0.3481
	(0.6055)	(0.5783)	(0.6182)	(0.6253)
rds_std	-0.1090	-0.1251	-0.1120	-0.1191
	(0.1015)	(0.0969)	(0.1034)	(0.1072)
mediate	0.4595**	0.4369**	0.4938**	0.2261
	(0.1677)	(0.1606)	(0.1741)	(0.1797)
lnedu	2.2349*	3.0113**	2.4818*	3.1393**
	(0.9860)	(0.9432)	(1.0009)	(1.0605)
_cons	-2.4759	1.1068	-2.6894	-7.4644
	(5.5140)	(5.2707)	(5.6366)	(5.7097)
N	580	579	551	500
R^2	0.452	0.400	0.454	0.484
adj. R^2	0.396	0.340	0.397	0.428
F	17.3250	14.0288	17.2532	16.8806
p	0.0000	0.0000	0.0000	0.0000

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Among the regressions: (1) represents the original regression results; (2) represents the regression results after replacing the dependent variable (the natural logarithm of the number of parties involved in collective labor dispute cases is used as the new dependent variable); (3) represents the regression results after excluding the year 2020. Considering that the special impact of the COVID-19 pandemic in 2020 may interfere with the research results, this paper adjusts the sample data and

re-conducts the regression analysis after excluding 2020 to test the robustness of the model results; (4) represents the regression results after excluding municipalities directly under the Central Government. This is because municipalities directly under the Central Government have particularities in terms of economic development level and policy environment, which may bias the overall regression results.

It can be seen that the results of each robustness test are basically consistent with the original regression results in terms of the significance of independent variables, the direction of coefficients, and the magnitude of coefficients, indicating that the regression model has strong robustness.

3.2.3 Regional Heterogeneity Analysis

Considering that differences in economic development level, policy environment, and social structure across regions may affect the regression results, this paper further groups the data by region to conduct heterogeneity analysis. Specifically, referring to the national provincial division method by Shen Xiaobo et al, the sample is divided into three groups: eastern, central, and western regions. Regression analysis is conducted for each group to explore the differences in the relationships between variables across regions, thereby further verifying the generalizability and regional differences of the research conclusions. The results of the regional heterogeneity analysis are shown in Table 6:

Table 6. Heterogeneity Analysis Results

	(1) Original Regression	(2) Eastern Region	(3) Central Region	(4) Western Region
law	-3.2537**	-2.0109	-6.0684*	-2.4840
	(1.1332)	(1.5194)	(2.3300)	(2.4328)
lngdp	1.0929***	1.8548***	1.0617	-0.5634
	(0.2275)	(0.3496)	(0.5370)	(0.5516)
unemploy_std	0.5616***	0.3446**	-0.2682	0.5261
	(0.1291)	(0.1242)	(0.6014)	(0.6399)
lnwage	-0.5846	-1.3096	0.0389	1.1817
	(0.6055)	(0.8218)	(1.2056)	(1.3163)
rds_std	-0.1090	-0.1672	0.8158*	-0.3780
	(0.1015)	(0.1021)	(0.3877)	(0.8842)
mediate	0.4595**	0.4883	0.0179	0.7861*
	(0.1677)	(0.2550)	(0.3127)	(0.3124)
lnedu	2.2349*	0.2244	5.3008*	-0.8158
	(0.9860)	(1.6614)	(2.3640)	(1.7676)
_cons	-2.4759	1.9006	-14.1066	-0.3745
	(5.5140)	(8.0832)	(11.4092)	(11.7523)
N	580	220	160	200
R^2	0.452	0.405	0.756	0.390
adj. R^2	0.396	0.292	0.695	0.264
F	17.3250	5.0099	15.7507	4.2162
p	0.0000	0.0000	0.0000	0.0000

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

(1) Legal Environment. The negative impact of the legal variable (revision of the Labor Contract Law) on collective labor disputes is significant in the overall sample but varies across regions. Specifically, this variable shows a significantly negative correlation in the central region, with an impact intensity much higher than the national average. In contrast, the impact of this variable is not significant in the eastern and western regions.

(2) Economic Environment. The significant positive impact of Gross Regional Product (lngdp) on the number of collective labor dispute cases is mainly reflected in the eastern region, with an impact intensity much higher than the national average. In the central and western regions, the impact of lngdp is not significant. The average wage of employed persons (lnwage) shows no statistically significant impact on the number of collective labor dispute cases in all regional samples, which is consistent with the overall regression results. This further supports the conclusion that this variable fails to capture a direct and significant impact on the total number of collective labor disputes in this model, even in regional sub-samples.

(3) Political Environment. The number of unemployment insurance participants (unemploy_std) shows a significant positive impact on collective labor disputes in the overall sample. At the regional level, this positive relationship is only significant in the eastern region, but its impact intensity is lower than the national average. In the central and western regions, the impact of the number of unemployment insurance participants is not significant, which may be related to regional differences in the development level of the social security system, workers' awareness of rights protection, and employment structure. The success rate of labor dispute mediation (mediate) shows a significant positive correlation with the number of collective labor disputes in the overall sample. In the regional analysis, this positive relationship is mainly driven by the western region, with an impact intensity higher than the national average. In contrast, the impact of the mediation success rate is not significant in the eastern and central regions, which may be related to differences in the maturity of mediation mechanisms and workers' awareness and usage habits of mediation channels across these regions.

(4) Socio-Cultural Environment. The significant positive impact of average years of schooling (lnedu) on collective labor disputes is mainly reflected in the central region, with an impact intensity much higher than the national average. This indicates that in the central region, the improvement of workers' educational level may more significantly enhance their awareness of rights, legal knowledge, and willingness to safeguard rights, thereby promoting more collective disputes to be resolved through formal channels.

(5) Technological Environment. The R&D full-time equivalent (rds_std) is not significant in the overall regression but shows notable differences in the regional heterogeneity analysis. This indicates that in the central region, the increase in investment in scientific and technological human capital is significantly associated with the growth in the number of collective labor dispute cases. In contrast, the impact of rds_std is not significant in the eastern and western regions.

4. Conclusions

The main conclusions of this study are summarized as follows:

First, from the overall national perspective, the "five-in-one" analytical framework for the labor relations environment exhibits significant explanatory power in

accounting for fluctuations in the number of collective labor disputes. In terms of the legal environment, the implementation of the revised Labor Contract Law significantly reduced the number of collective labor dispute cases. This indicates that a sound legal system and effective law enforcement are crucial guarantees for regulating the behaviors of labor and management and reducing group conflicts. In terms of the economic environment, there is a significantly positive correlation between the growth of Gross Regional Product (GRP) and the number of collective labor disputes, reflecting the contradictions that may arise from industrial structure adjustment and the complication of labor relations during economic development. However, the impact of the average wage of employed personnel on the number of collective labor disputes did not reach a statistically significant level. In terms of the political environment, the increase in the number of unemployment insurance participants significantly raised the number of collective labor dispute cases, which may be related to the enhanced awareness and ability of workers to safeguard their rights. Meanwhile, the improvement in the success rate of labor dispute mediation also showed a significantly positive correlation with the number of collective labor dispute cases—this may imply that a more effective mediation mechanism has, in turn, prompted more potential disputes to enter formal resolution channels, thereby increasing the degree of explicitness of cases. In terms of the socio-cultural environment, the improvement in the average years of schooling significantly increased the number of collective labor dispute cases. This may stem from the fact that workers with higher educational attainment possess stronger awareness of rights and legal knowledge and are more inclined to safeguard their legitimate rights and interests through formal channels. In terms of the technological environment, the R&D full-time equivalent did not exert a statistically significant impact on the number of collective labor disputes. Combined with the analysis in this paper, this may suggest that technological progress has not fundamentally transformed the existing production relations at the macro level, and the phenomena of exploitation and oppression in the labor field remain the core contradictions triggering collective labor disputes. Therefore, mere technological input has failed to significantly affect the total number of collective disputes.

Second, the regional heterogeneity analysis reveals the complexity and regional differences in the mechanism of action of various influencing factors: The inhibitory effect of the revised Labor Contract Law is particularly significant in the central region, while it is not significant in the eastern and western regions. This may indicate that the central region had more room for improvement in the standardization of labor employment prior to the law's revision, hence the more pronounced regulatory effect of the legal revision. The significant positive impact of GRP is mainly reflected in the eastern region, indicating that labor-management contradictions arising from rapid economic growth are more prominent in the east; this impact is not significant in the central and western regions. The positive effect of the number of unemployment insurance participants is also mainly manifested in the eastern region, suggesting that the enhancement of workers' awareness of rights protection in the eastern region is more closely linked to the improvement of the social security system. The R&D full-time equivalent is not significant at the national level but shows a significantly positive impact in the central region. This may reflect that the introduction of new technologies has caused certain impacts and contradictions on labor relations during the process of industrial transformation and technological application in the central region. The success rate of labor dispute mediation shows a significant positive correlation in the western region, indicating that the improvement of the mediation mechanism in the

western region may have more effectively promoted the explicitness of disputes. The positive impact of the average years of schooling is particularly prominent in the central region, which may indicate that the improvement of workers' educational level in the central region has the most significant promoting effect on their awareness of rights and rights-protection actions. These heterogeneous findings emphasize that when understanding and addressing collective labor disputes, full consideration must be given to the economic development stage, social structure, institutional environment, and cultural characteristics of different regions.

5. Research Limitations

Despite the theoretical and methodological contributions of this study, it still has certain limitations that need to be acknowledged and addressed in future research:

(1) Sample and Data Limitations. First, the sample excludes the Tibet Autonomous Region and the Xinjiang Production and Construction Corps due to data availability issues, which may lead to incomplete coverage of China's regional labor relations characteristics, especially the unique labor dispute patterns in ethnic minority and border areas. Second, the selection of indicators is relatively simplified: for example, the technological environment is only measured by R&D full-time equivalent, failing to cover indicators such as digitalization penetration rate, platform employment scale, and technological substitution intensity, which may underestimate the impact of new technological changes on labor relations. Third, the study uses macro-provincial panel data, lacking micro-level data such as enterprise surveys or employee questionnaires, making it difficult to deeply explore the internal mechanism of factors affecting collective labor disputes at the enterprise and individual levels.

(2) Model and Method Limitations. Although the two-way fixed effects model is adopted to control for regional and time fixed effects, the study does not fully consider the interaction effects between variables (e.g., the joint impact of legal environment and educational level on workers' rights-protection behavior) and potential reverse causality (e.g., the surge in collective labor disputes may prompt local governments to strengthen labor law enforcement or adjust social security policies). In addition, the study focuses on linear relationships between variables, and there may be non-linear correlations (such as the "inverted U-shaped" relationship between technological progress and disputes in specific regions) that have not been fully explored.

(3) Research Scope Limitations. The study focuses on the quantitative impact of environmental factors on collective labor disputes but pays less attention to the qualitative characteristics of disputes, such as the severity of conflicts, resolution efficiency, and long-term impacts on labor relations. Moreover, it does not distinguish between different types of collective labor disputes (e.g., wage disputes, social insurance disputes, working condition disputes), leading to the inability to identify the heterogeneous influencing factors of specific types of disputes.

6. Future Research Directions

In response to the aforementioned limitations and generalizability issues, future research can be expanded in the following directions:

(1) Expand Sample and Enrich Data. First, efforts should be made to collect data from ethnic minority and border areas to supplement the sample and improve the comprehensiveness of regional coverage. Second, refine indicator systems: introduce micro-level data such as enterprise ownership type, trade union participation rate, and workers' rights-protection awareness; add indicators such as digital employment proportion and technological substitution rate to measure the technological environment more accurately. Third, combine macro data with micro surveys, such as conducting paired surveys of enterprises and employees, to construct a multi-level data set and deeply analyze the transmission mechanism between macro environmental factors and micro labor dispute behaviors.

(2) Optimize Research Methods and Models. First, introduce interaction terms into the model to explore the joint impact of multiple factors on collective labor disputes; use causal identification methods such as instrumental variables and difference-in-differences (DID) to solve the problem of reverse causality and improve the credibility of causal inferences. Second, conduct subgroup analysis by classifying collective labor disputes into different types (e.g., wage disputes, working condition disputes) and different industries, to identify the heterogeneous influencing factors of specific disputes. Third, explore non-linear relationships between variables through threshold regression, polynomial regression, and other methods, and further verify the regional differences in the impact of factors.

(3) Expand Research Scope and Depth. First, strengthen qualitative research on collective labor disputes, such as analyzing the evolution process, conflict points, and resolution paths of typical cases, to complement the limitations of quantitative research. Second, focus on new employment forms and new types of labor disputes, such as collective disputes involving platform workers, gig workers, and flexible employees, and explore the adaptability of traditional labor dispute resolution mechanisms in the digital economy era. Third, carry out cross-country comparative research, select countries with different institutional backgrounds (e.g., developed countries in Europe and America, emerging economies in Asia) for comparative analysis, and explore the universal and specific factors affecting collective labor disputes in different institutional contexts. Fourth, pay attention to the long-term effects of collective labor disputes, such as the impact on enterprise innovation, regional economic development, and social stability, to provide more comprehensive policy implications.

References

[1] Matthias Fahn,Reciprocity in Dynamic Employment Relationships. Management Science.2003;69(10):5816-5829,doi:10.1287/mnsc.2022.4657.Sherer P.D., Rogovsky N. & Wright N.What Drives Employment Relationships in Taxicab Organizations? Linking Agency to Firm Capabilities and Strategic Opportunities. Organization Science.1998;34-48,doi:10.1287/orsc.9.1.34.

[2] Buchele, R., & Christiansen, J.Employment and Productivity Growth in Europe and North America: The Impact of Labor Market Institutions. International Review of Applied Economics.1999;13(3), 313,doi:10.1080/026921799101571.

[3] Vieriu, E., & Vieriu, D. General Aspects Regarding the Evolution of the Regulations of the Collective Labor Conflicts in Romania. Jus et Civitas.2017;IV (LXVIII)(1), 67–74.

[4] Hui C., Lee C., & Rousseau D.M. Employment Relationships in China: Do Workers Relate to the Organization Or to People? Organization Science.2004;15, 232-240,doi:10.1287/orsc.1030.0050.

[5] GUŢă, A. J. The Role of Effective Communication in the Prevention of Labor Conflicts within an Institution of Local Public Administration. Review of International Comparative Management/Revista de Management Comparat International.2024;25(5), 884–896,doi:10.24818/rmci.2024.5.884.

[6] Volosevici, D. Alternative Dispute Resolution and Individual Labour Conflicts: A Different Way of Managing Organizational Disputes. Jus et Civitas.2021;2, 56–64,doi:10.51865/jetc.2021.02.08.
[7] Simone G. Labour Conflicts in the Digital Age-a Comparative Perspectiveby Donatelladella Porta, Riccardo Emiliochesta and Lorenzocini. Isbn: 978-1529228243. Price £80. British Journal of Industrial Relations.2023;61, 986-988,doi:10.1111/bjir.12769.
[8] WANG Zhiwei, ZHAO Jilun, YU Guilan. The Growth and Causes of Labor Dispute Cases in China: An Empirical Analysis Based on Relevant Data from 1991 to 2016. Journal of Shandong University (Philosophy and Social Sciences).2020;(5): 64-73,doi:10.19836/j.cnki.37-1100/c.2020.05.007.
[9] KONG Lingming.An Empirical Study on Collective Labor Disputes: Based on Original Data of 3,110 Disputes in City D of a Certain Province in the Past Decade. Tribune of Social Sciences.2016;(10): 214-223,doi:10.14185/j.cnki.issn1008-2026.2016.10.023.
[10] GAO Hongyan.An Empirical Study on Influencing Factors of Labor Dispute Growth in China's Economic Transition Period. Taxation and Economy.2012;(3): 35-42,doi:CNKI:SUN:SWYJ.0.2012-03-008.
[11] ZHAN Yubo, CHEN Qiang.Foreign Trade, Processing Trade and Labor Disputes: Evidence from China's Inter-provincial Panel Data (1998-2012). Labor Economics Research.2018;6(1): 81-106,doi:CNKI:SUN:LDJJ.0.2018-01-005.
[12] ZHANG Shaoping.An Economic Analysis of Influencing Factors of Labor Disputes. Journal of China Institute of Industrial Relations.2012;26(3): 37-41,doi:10.3969/j.issn.1673-2375.2012.03.007.
[13] WANG Yang.Analysis of Influencing Factors of Labor Disputes in China During the Transition Period. Research on Economics and Management.2013;(4): 71-80,doi:10.3969/j.issn.1000-7636.2013.04.008.
[14] CHENG Xiao, LI Jing.Labor Disputes, Capital Deepening and Labor Income Share: Influence Mechanism and Chinese Experience. Western Forum.2021;31(1): 48-58,doi:10.3969/j.issn.1674-8131.2021.01.005.
[15] HU Lei.Characteristics and Implications of the Incidence of Labor Dispute Cases in China: An Analysis Based on Panel Data (1991-2011). Journal of Wuhan University of Technology (Social Sciences Edition).2014;27(5): 801-805,doi:10.3963/j.issn.1671-6477.2014.05.014.
[16] YU Qin, ZHUANG Wenjia. Labor Legislation, Local Mediation and Dispute Resolution in the 40 Years of Reform and Opening-up: An Empirical Analysis Based on the Winning Rate of Labor Disputes. Journal of Sun Yat-sen University (Social Science Edition).2018;58(3): 171-177,doi:CNKI:SUN:ZSDS.0.2018-03-019.
[17] CHEN Tianxue, ZHAO Man.Forecasting Methods for the Number of Labor Dispute Cases Under the Constraint of Labor Law. Statistics & Decision.2014;(1): 86-88,doi:CNKI:SUN:TJJC.0.2014-01-026.
[18] ZHUANG Wenjia.Can "Mediation First" Alleviate Collective Labor Disputes? An Empirical Test Based on Inter-provincial Panel Data (1999-2011). Sociological Studies.2013;28(5): 145-171+244-245,doi:10.19934/j.cnki.shxyj.2013.05.007.
[19] LUO Yan, LIN Qiulan.An Empirical Analysis of Collective Labor Disputes: Based on Data from 31 Provinces, Municipalities and Autonomous Regions in China. Journal of China Institute of Industrial Relations.2011;25(2): 50-55,doi:CNKI:SUN:GHLJ.0.2011-02-012.
[20] CHANG Cheng.Characteristics of Collective Labor Disputes in China and an Empirical Study on the Protection of Workers' Rights and Interests: Based on a Survey of 106 Collective Labor Disputes in Xicheng District, Beijing. Shandong Trade Unions' Tribune.2019;25(1): 72-77,doi:10.3969/j.issn.1008-6153.2019.01.011.
[21] KONG Lingming.An Empirical Study on Labor Disputes Caused by Defaulting on Wages and Fleeing: Based on Empirical Data from City D, Guangdong Province (2004-2016). Journal of Wenzhou University (Social Science Edition).2018;31(3): 27-34,doi:CNKI:SUN:WZDX.0.2018-03-005.
[22] CHEN Chong, HE Mengxing.A Differential Analysis of Influencing Factors of Labor Disputes in Northwest China. Journal of Xidian University (Social Sciences Edition).2016;26(2): 47-54,doi:10.16348/j.cnki.cn61-1336/c.2016.02.006.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261137

Multi-Task Surface State Perception Enhanced by Self-encoder-decoder Features

Quan GAN [a,1] Qidan ZHU [a] Jing LI [b] and Yang TIAN [c]
[a] *College of Intelligent Systems Science and Engineering, Harbin Engineering University, Harbin, 150001, China*
[b] *School of Automation, Chengdu University of Information Technology, Chengdu, 610225, China*
[c] *Guangxi University, Nanning, 53004, China*

Abstract. In multi-task water surface perception models, conflicting task objectives often lead to degraded generalization of shared features. To address this issue, we propose a novel multi-task perception framework based on self-encoder-decoder feature enhancement, termed SEDFNet. The model takes RGB images and 3D radar point clouds as dual-modal inputs, and feeds the extracted shared features into three parallel task branches: object detection, semantic segmentation, and drivable area segmentation. To further enhance the discriminability and robustness of shared features, this paper introduces an auxiliary self-supervised reconstruction task, where the decoder reconstructs the original input image from the shared features, and the mean squared error between the reconstructed image and the original image is used as an additional constraint. This method enables the backbone network to implicitly learn feature representations with better information integrity and task independence during training, thereby improving the efficiency of feature reuse among multiple tasks. Experiments on the WaterScenes dataset demonstrate that the proposed self-encoder-decoder feature enhancement method effectively strengthens shared feature representations, and the constructed model achieves competitive performance across multiple tasks.

Keywords. Multi-task, Surface state perception, Self-encoder-decoder, Feature enhancement

1. Introduction

In recent years, the environmental perception technology of Unmanned Surface Vehicles (USVs) has made considerable progress in the field of autonomous driving, among which multi-sensor fusion and multi-task learning have become two key directions driving the improvement of system performance[1, 2]. Multi-sensor fusion achieves complementary advantages between sensors by collaboratively processing heterogeneous data such as cameras, lidar, and radar, which significantly enhances the perception robustness and accuracy of the system in complex surface environments such as fog, rain, and waves[3]. At

[1]Corresponding Author: Quan Gan; E-mail: gquan@hrbeu.edu.cn.

the same time, to support efficient autonomous navigation and path planning, USVs need to have the ability to perform multiple types of perception tasks simultaneously, such as object detection, semantic segmentation, and drivable area segmentation. By sharing network structures and feature representations, multi-task learning not only expands the system's global understanding of scenes but also improves the information utilization efficiency between tasks, thereby achieving better overall performance under resource-constrained conditions. Therefore, developing an efficient and stable multi-task learning framework is of great significance for enhancing the comprehensive environmental perception and decision-making capabilities of USVs [4].

In the typical architecture of multi-task learning models, the shared backbone network is responsible for extracting general semantic features for downstream tasks, and each task branch performs specific outputs based on these features [5, 6]. However, this architecture often faces the problem of degraded generalization ability of shared features caused by inconsistent learning objectives among tasks, thereby weakening the generalization ability of shared features and limiting the overall performance of the model. To alleviate feature interference of the shared backbone network under multi-task settings and enhance its feature representation capability, this study proposes a multi-task water surface state perception method based on self-encoder-decoder feature enhancement. This method uses RGB images and 3D radar point clouds as dual-modal inputs: RGB images extract basic semantic features through the shared backbone network; 3D point clouds are compressed and projected into 2D radar data for preliminary feature extraction. The obtained basic semantic features are sent to task branches such as object detection, semantic segmentation, and drivable area segmentation. In the object detection branch, image features and radar features are fused to improve detection robustness. To further strengthen the feature learning of the shared backbone network, this paper introduces a self-decoding auxiliary task: input the extracted basic semantic features into the decoder to reconstruct the original input image, and use the mean squared error between the reconstructed image and the original RGB image as the auxiliary loss. Through this self-supervised reconstruction task, the backbone network is constrained to learn more discriminative and robust visual representations, thereby achieving implicit feature enhancement in the feature extraction stage. Experiments show that this self-encoding auxiliary mechanism effectively reduces the gradient interference between multiple tasks and significantly improves the comprehensive performance of the surface state perception model in multiple tasks.

In summary, the main contributions of this paper are twofold:

1.This paper proposes a novel multi-task water surface state perception framework based on self-encoder-decoder feature enhancement.

2.The proposed multi-task water surface perception model SEDFNet is evaluated on the WaterScenes dataset and achieves highly competitive results.

2. Methodology

The multi-task surface state perception model SEDFNet proposed in this paper adopts a dual-branch architecture design, as shown in Figure 1. Its core lies in integrating three frameworks: encoder-decoder, semantic segmentation, and object detection, to achieve multi-modal perception and multi-task collaborative analysis of surface scenes. The

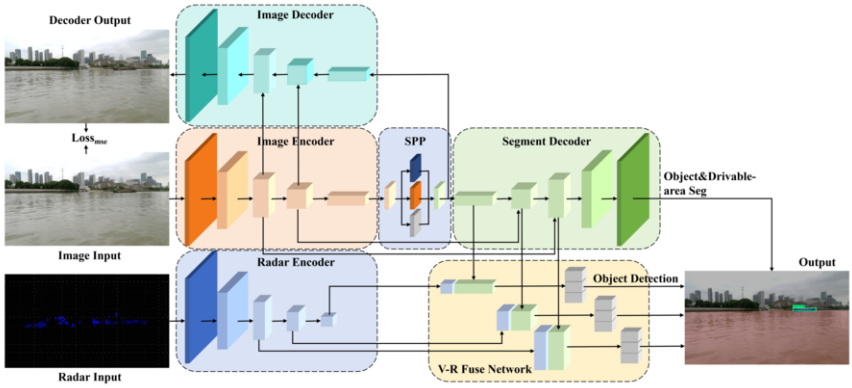


Figure 1. Overall architecture diagram of the multi-task surface state perception model SEDFNet.

model includes two independent input branches, each equipped with a dedicated encoder, which can process different modal data separately. The main body of the SEDFNet model consists of three parts: a feature enhancement module based on a self-encoding-decoding structure, an object detection module based on multi-modal fusion, and a semantic segmentation module based on an encoding-decoding structure, which overall realizes efficient multi-level feature extraction and task collaboration.

2.1. Feature Enhancement Module Based on Self-Encoding-Decoding Structure

In multi-task learning models, the conventional architecture usually adopts a design that shares the backbone network and branches only at different task heads. While balancing efficiency and simplicity, this architecture also makes full use of the shared representation of low-level features. However, this sharing mechanism also has obvious limitations: when the correlation between tasks is low, or conflicts occur, shared features may lead to performance degradation, that is, negative transfer. Therefore, introducing self-supervised feature enhancement methods to learn feature representations with higher information integrity and task irrelevance is particularly important for alleviating negative transfer and improving the model's feature representation ability.

For an image $I \in \mathbb{R}^{3\times H\times W}$ of size $H \times W$, input I into the encoder CNN_{encoderI} with CNN as the backbone network:

$$f_{I_1} = CNN_{encoderI}(I) \in \mathbb{R}^{C\times h\times w} \tag{1}$$

Among them, f_{I_1} is the encoded feature, C is the number of channels after encoding, $h \times w$ is the size of the feature map after encoding, and intermediate results are saved during the encoding process to obtain feature maps f_{I_2} and f_{I_3} of different resolutions. Then, input f_{I_1} into SPP for multi-scale feature enhancement:

$$f_{spp} = SPP(f_{I_1}) \in \mathbb{R}^{C\times h\times w} \tag{2}$$

Among them, f_{spp} is the result of multi-scale feature enhancement, C is the number of channels, and $h \times w$ is the size of the feature map. Then input f_{spp} into the reconstruction branch, and reconstruct the input data through the decoder:

$$I^{'} = CNN_{decoderI}(f_{spp}) \in \mathbb{R}^{3\times H\times W} \tag{3}$$

Among them, $I^{'}$ is the reconstructed data, and the reconstruction error is calculated with the input data through the mean squared error loss function.

2.2. Object Detection Module Based on Multi-Modal Fusion

SEDFNet adopts a dual-branch structure, which receives two different modal data of image and radar as inputs, respectively. In the object detection process, the model improves detection performance by extracting features of the two modalities separately and fusing them effectively. For an image $I \in \mathbb{R}^{3\times H\times W}$ of size $H \times W$, input I into the encoder CNN_{encoderI} with CNN as the backbone network. After encoding, SPP is used for multi-scale feature enhancement to obtain, as shown in equations (1) and (2). A series of feature maps with gradually increasing resolutions is generated by combining convolution and upsampling of f_{spp}, and fused with the corresponding encoded feature maps:

$$f_{S_1} = Conv(Cat(UP_Conv(f_{spp}), f_{I_1})) \in \mathbb{R}^{C_1\times h_1\times w_1} \tag{4}$$

$$f_{S_2} = Conv(Cat(UP_Conv(f_{S_1}), f_{I_2})) \in \mathbb{R}^{C_2\times h_2\times w_2} \tag{5}$$

$$f_{S_3} = Conv(Cat(UP_Conv(f_{S_2}), f_{I_3})) \in \mathbb{R}^{C_3\times h_3\times w_3} \tag{6}$$

Among them, f_{S_1}, f_{S_2} and f_{S_3} are feature maps of different resolutions.

The radar data $R \in \mathbb{R}^{3\times H\times W}$ is feature-encoded by the radar encoder CNN_{encoderR}:

$$f_{R_1} = CNN_{encoderR}(R) \in \mathbb{R}^{C'\times h\times w} \tag{7}$$

Among them, f_{R_1} is the number of channels after encoding, C' is the size of the feature map after encoding, which is consistent with the size of the feature map after image data encoding. During the encoding process, feature maps f_{R_2} and f_{R_3} those of different encoding stages are generated. The three radar encoding results of different resolutions f_{R_1}, f_{R_2} and f_{R_3} are fused with the image pyramids c respectively, and feature enhancement is performed through the attention mechanism:

$$f_{head_1} = SE(Cat(f_{S_1}, f_{R_1})) \in \mathbb{R}^{C_1'\times h_1\times w_1} \tag{8}$$

$$f_{head_2} = SE(Cat(f_{S_2}, f_{R_2})) \in \mathbb{R}^{C_2'\times h_2\times w_2} \tag{9}$$

$$f_{head_3} = SE(Cat(f_{S_3}, f_{R_3})) \in \mathbb{R}^{C_3'\times h_3\times w_3} \tag{10}$$

Among them, f_{head_n}, $n = 1,2,3$ is the detection head, C_n', $n = 1,2,3$ is the number of channels after fusion. Finally, the detection head is sent to the object detector for detection and output results:

$$f_{out_d} = Header(f_{head_1}, f_{head_2}, f_{head_3}) \tag{11}$$

Among them, f_{out_d} is the final output of object detection.



Figure 2. Some image examples of the WaterScenes dataset.

2.3. Semantic Segmentation Module Based on Encoding-Decoding Structure

When SEDFNet performs semantic segmentation, only image data is used as input. For an image of size $H \times W$, input I into the encoder with CNN as the backbone network. After encoding, feature maps f_{I_1}, f_{I_2} and f_{I_3} of the encoding stage are obtained. Then, SPP is used for multi-scale feature enhancement to obtain f_{spp}, as shown in equations (1) and (2). Then, a series of feature images with gradually increasing resolutions is generated by combining convolution and upsampling of f_{spp}, which is also a decoding process, as shown in equations (4), (5), and (6), to obtain feature maps f_{S_1}, f_{S_2} and f_{S_3} of different resolutions. Then, continue the decoding operation of f_{S_3} to obtain the final semantic output:

$$f_{out_s} = Decoder(f_{S_3}) \tag{12}$$

Among them, f_{out_s} is the final output of semantic segmentation.

3. Experiment

3.1. Dataset and Judgment Criteria

The experiments in this study were verified using the WaterScenes dataset[7], which is the first multi-task 4D millimeter-wave radar-camera fusion dataset for surface scenes. It provides multi-sensor synchronous data, including 4D millimeter-wave radar, monocular camera, GPS, and IMU, supporting six major perception tasks: object detection, instance segmentation, semantic segmentation, drivable area segmentation, water shoreline seg-

mentation, and panoramic perception. This study only completes three tasks: object detection, semantic segmentation, and drivable area segmentation. Data collection covers different times (day, dusk, night), lighting conditions (normal, dim, strong light), weather (sunny, cloudy, rainy, snowy), and waterway environments (rivers, lakes, canals, moats), including a total of 54,120 sets of RGB images, radar point clouds, GPS, and IMU data, covering more than 200,000 surface targets. The image resolution of the dataset is 1920×1080, and the target categories include piers, buoys, sailors, ferries, ships, vessels, and kayaks; among them, buoys and piers are typical static obstacles that unmanned ships need to avoid during navigation, while ships, ferries, vessels, and kayaks represent common waterborne dynamic vehicles. "Sailors" specifically refer to personnel on these vehicles. Some image examples of the dataset are shown in Figure 2, and the number of image samples for each category is shown in Table 1. To comprehensively and objectively evaluate the performance of the model, for the object detection task, this study uses mAP50, mAP50-95, and AR50 as evaluation metrics; for the target semantic segmentation and drivable area segmentation tasks, mIoU is selected as the performance metric.

Table 1. Number of image data of each category (The total number of frames represents the total number of images, and the number of frames for each category represents the number of images containing the corresponding target. A single image may contain multiple targets, so it is counted in multiple categories.)

	Total	Pier	Buoy	Sailor	Ship	Boat	Vessel	Kayak
Frames	54,120	25,787	3,769	3,613	19,776	9,106	9,362	366
Objects	202,807	121,827	16,538	8,036	34,121	10,819	11,092	374

3.2. Implementation Rules

In the model experiment, this study adopted the cosine learning rate scheduler and SGD optimizer, with an initial learning rate of 0.03 and a batch size of 32. The experiment was carried out on the NVIDIA RTX 3090Ti GPU platform with 24GB of dedicated video memory. To save GPU memory, the training framework also supports mixed-precision training. In addition, an exponential moving average is used to alleviate model overfitting, thereby improving generalization ability and stability. In the data preprocessing stage, the input data is normalized and uniformly scaled to a resolution of 320320. The entire training process lasts for 100 epochs.

3.3. Experimental results

The multi-task fusion model SEDFNet proposed in this paper has achieved highly competitive results on the WaterScenes dataset, as shown in Table 2, especially showing advantages in the collaborative performance of object detection and semantic segmentation. As a multi-task model fusing visible light and radar information (C+R), SEDFNet processes three tasks simultaneously: object detection, semantic segmentation, and drivable area segmentation. In the systematic comparison with other single-task and multi-task models, its overall performance ranks among the top, and it leads in multiple key indicators.

In terms of object detection, SEDFNet achieves 79.5 in the mAP50 metric, which is leading among all participating multi-task models, an increase of 0.1 compared with

the MV-GDF-X-PN-S2 model[8]. This indicates that SEDFNet has excellent recognition accuracy under loose detection thresholds. In the mAP50-95 metric that measures comprehensive detection accuracy, SEDFNet achieves 43.6, which is slightly lower than 45.0 of MV-GDF-X-PN-S2[8], but still significantly better than the ASY-VRNet model[9], YOLOP model[10], the HybridNets model[11], and some single-stage detection models. In addition, its AR50-95 is 48.1, and the recall ability is also at a medium-to-high level.

Table 2. Performance comparison of SEDFNet with other models on the WaterScenes dataset (TN: task number, OD: object detection, SS: semantic segmentation, $mIoU_t$: mIoU of targets, $mIoU_d$: mIoU of drivable area)

Models	Sensors	TN	OD			SS	
			mAP50-95	mAP50	AR50-95	$mIoU_t$	$mIoU_d$
Single-Task Models							
Object Detection							
YOLOv8-Nano[12]	C	1	41.9	71.8	44.0	-	-
YOLOv7-Tiny[13]	C	1	37.3	65.9	43.7	-	-
YOLOX-Tiny[14]	C	1	39.4	68.0	43.0	-	-
YOLOv4-Tiny[15]	C	1	13.1	36.3	20.2	-	-
CRFNet[16]	C+R	1	41.8	71.2	44.5	-	-
Semantic Segmentation							
Segformer-B0[17]	C	2	-	-	-	73.5	99.4
PSPNet[18]	C	2	-	-	-	69.4	99.0
Multi-Task Models							
YOLOP[10]	C	2	37.9	68.9	43.5	-	99.0
HybridNets[11]	C	2	39.1	69.8	44.2	-	98.8
Efficient-VRNet-N[2]	C+R	3	42.0	73.5	44.2	72.1	99.3
MV-GDF-X-PN-S0[8]	C+R	3	41.5	71.3	45.6	70.6	99.5
MV-GDF-X-PN-S1[8]	C+R	3	43.1	75.8	47.2	73.2	99.5
MV-GDF-X-PN-S2[8]	C+R	3	45.0	79.4	48.8	73.8	99.6
ASY-VRNet[9]	C+R	3	42.8	-	46.3	74.7	99.6
SEDFNet	C+R	3	43.6	79.5	48.1	76.9	98.7

In terms of semantic segmentation, SEDFNet performs extremely well, with an mIoUt metric as high as 76.9, far exceeding other multi-task models. This fully proves that under the multi-task learning framework, SEDFNet not only does not sacrifice segmentation accuracy but also significantly enhances the ability of scene understanding and pixel-level classification through feature sharing. Although the mIoUd of SEDFNet in drivable area segmentation is 98.7, which is slightly lower than some models, it still maintains an extremely high level, and the overall segmentation performance is comprehensive and strong.

In summary, SEDFNet has successfully achieved performance balance and mutual improvement among multiple tasks, especially maintaining advanced object detection capabilities while improving semantic segmentation accuracy. This paper shows some examples of surface perception visualization results in Figure 3.

This paper presents and analyzes some failure cases of the SEDFNet model on the WaterScenes dataset in Figure 4. As shown in the first group of cases in Figure 4, under extremely dark lighting conditions, the model mistakenly identifies reflections in the water as a dock. The second group of cases shows that when a ship is loaded with cargo,



Figure 3. Example of visual results of water surface state perception using the SEDFNet model on the WaterScenes dataset.

the model fails to completely segment the hull. In the third group of cases, under heavy fog weather, the model fails to detect small distant target ships. The fourth group of cases indicates that when a ship is partially submerged in water and occluded, the model cannot successfully detect the ship. In summary, the SEDFNet model still has certain limitations when dealing with scenarios such as extremely low illumination, heavy occlusion, heavy fog, and small-target detection.

3.4. Visualization of Shared Feature Heatmaps

In this section, a heatmap visualization experiment of shared features is conducted on the WaterScenes dataset to intuitively demonstrate the effectiveness of the proposed self-decoder-based feature enhancement method. We adopt the Grad-CAM method to visualize the output of the SPP module (the output of the SPP module is the last layer of shared features), and the results are shown in Figure 5. In the heatmaps, the red areas indicate the parts focused on by the model, while the blue areas correspond to the non-focused areas.

Figure 5 presents the SPP feature heatmaps under four different settings: the first column shows the SPP feature heatmaps of the object detection task only (Det); the second column shows the SPP feature heatmaps of the semantic segmentation task only (Seg); the third column shows the SPP feature heatmaps of the joint task including both object detection and semantic segmentation (Det+Seg); the fourth column shows the SPP feature heatmaps of the task incorporating the self-encoder-decoder task on the basis of object detection and semantic segmentation (Det+Seg+SED).

As can be observed from the comparison, in Groups 2, 3, and 8, the Det+Seg task exhibits the phenomenon of shared feature degradation, while the Det+Seg+SED with

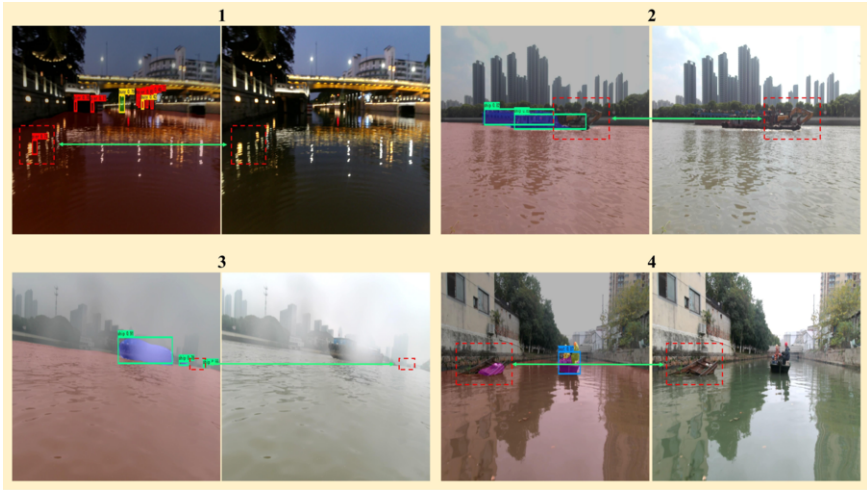


Figure 4. Visualization Examples of Failure Cases for the SEDFNet Model in Water Surface State Perception on the WaterScenes Dataset.

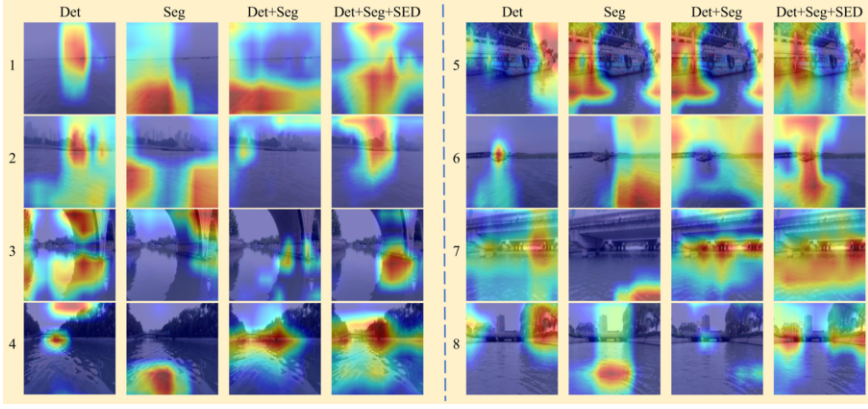


Figure 5. Visualization of shared feature heatmaps under different tasks (where Det: Object Detection; Seg: Semantic Segmentation; SED: self-encoder-decoder).

the introduction of the self-encoder-decoder task effectively alleviates this problem. In addition, in Groups 1, 4, 5, 6, and 7, Det+Seg+SED shows a wider range of feature-focused areas compared with Det+Seg, which further verifies the advantage of the proposed method in enhancing the expression of shared features.

3.5. Ablation experiment

In this section, ablation experiments are conducted on the WaterScenes dataset to further verify the effectiveness of each component of the model. For the object detection task, mAP50, mAP50-95, and AR50 are used as evaluation metrics; for the target semantic

segmentation and drivable area segmentation tasks, mIoU is selected as the performance metric.

Ablation of the self-encoder-decoder. When exploring the role of the self-encoder-decoder in surface state perception tasks, we compared the performance with and without the self-encoder-decoder, as shown in Table 3. The ablation study shows that the introduction of the self-encoder-decoder module has improved the model's performance in multiple metrics. Specifically, while keeping the sensor configuration and TN number unchanged, the model with the self-decoder increased mAP50-95 from 42.8 to 43.6, mAP50 from 77.5 to 79.5, and AR50-95 from 46.8 to 48.1 in the object detection task. In the semantic segmentation task, mIoUt has significantly increased from 67.13 to 76.9, and mIoUd of drivable area segmentation has also slightly increased from 98.6 to 98.7. The results show that the self-encoder-decoder effectively enhances the feature extraction ability of the model in multi-task perception, especially in the segmentation task.

Table 3. Ablation study of the self-encoder-decoder

Models	Sensors	TN	OD			SS	
			mAP50-95	mAP50	AR50-95	$mIoU_t$	$mIoU_d$
Without self-encoder-decoder	C+R	3	42.8	77.5	46.8	67.13	98.6
With self-encoder-decoder	C+R	3	43.6	79.5	48.1	76.9	98.7

Ablation of the number of tasks. To explore the impact of different numbers of tasks on model performance, the results are shown in Table 4. The ablation study shows that the multi-task joint learning model performs well in balancing various perception tasks. Specifically, the single object detection task (TN=1) performs well in detection metrics, with mAP50-95 reaching 43.7 and mAP50 reaching 79.1; when only two segmentation tasks are performed (TN=2), the segmentation results are better, with mIoU reaching 77.16 and mIoUd reaching 98.75. When the three tasks are trained jointly (TN=3), the model can maintain performance similar to or even better than that of single tasks: the object detection metrics mAP50-95 is 43.6, mAP50 is 79.5, which is slightly higher than that of the single detection task; the semantic segmentation metric mIoUt is 76.9, and the drivable area metric mIoUd is 98.7, which are also higher than those of the dual-task model. This indicates that the multi-task framework proposed in this paper has good collaboration and stability when integrating object detection and two segmentation tasks, and does not cause significant performance degradation due to the increase in tasks.

Table 4. Ablation study of the number of tasks(OD: Object Detection; SS+DAS: Semantic Segmentation + Drivable Area Segmentation; OD+SS+DAS: Object Detection + Semantic Segmentation + Drivable Area Segmentation)

Models	Sensors	TN	OD			SS	
			mAP50-95	mAP50	AR50-95	$mIoU_t$	$mIoU_d$
OD	C+R	1	43.7	79.1	47.7	-	-
SS+DAS	C+R	2	-	-	-	77.16	98.75
OD+SS+DAS	C+R	3	43.6	79.5	48.1	76.9	98.7

Ablation of input modalities. To explore the impact of input modalities on model performance, as shown in Table 5. The ablation study shows that the perceptual input fusing camera and radar can bring consistent performance improvements to the multi-

task model. With a fixed number of tasks, compared with using only camera data, the model combining camera and radar has improved in all five evaluation metrics: the object detection task's mAP50-95 increased from 43.1 to 43.6, mAP50 from 78.5 to 79.5, and AR50-95 from 47.2 to 48.1; the semantic segmentation task's mIoUt increased from 75.7 to 76.9, and mIoUd from 98.6 to 98.7. This indicates that the introduction of radar information effectively supplements the perceptual ability of camera data and enhances the overall performance of the model in multi-task scenarios.

Table 5. Ablation study of input modalities

Sensors	TN	OD			SS	
		mAP50-95	mAP50	AR50-95	$mIoU_t$	$mIoU_d$
C	3	43.1	78.5	47.2	75.7	98.6
C+R	3	43.6	79.5	48.1	76.9	98.7

4. Conclusion

To address the feature degradation problem caused by the discrepancy among task objectives in multi-task surface water perception models, this study proposes a Self-Encoding-Decoding Feature enhancement Network (SEDFNet), a multi-task perception method based on auto-encoder-decoder feature enhancement. This method uses RGB images and 3D radar point cloud data as inputs to extract shared features, and processes object detection, semantic segmentation, and drivable area segmentation tasks in parallel. Furthermore, a self-supervised reconstruction task is introduced. Specifically, the input image is reconstructed by decoding the shared features, and the reconstruction loss is leveraged to constrain the backbone network to learn feature representations with stronger information integrity and task-agnostic properties, thus improving the efficiency of feature reuse. Experiments on the WaterScenes dataset demonstrate that the proposed method can effectively enhance the discriminative power of shared features, enabling the model to achieve competitive performance across all tasks.

Although SEDFNet has been validated to be effective in alleviating the feature degradation problem on the WaterScenes dataset, there remain several promising directions for in-depth exploration in the current work. First, the underlying mechanism of the self-supervised reconstruction task's impact on feature learning has not yet been fully elucidated. In future work, we plan to introduce interpretability analysis methods to provide more guiding theoretical evidence for network architecture design. Second, the current model is mainly validated on a single dataset, and we will subsequently conduct experiments on more diverse water surface scene datasets. Furthermore, with the increasingly stringent real-time requirements for unmanned surface systems, how to reduce the computational overhead of the model via knowledge distillation or lightweight network design while maintaining its feature enhancement capability will also be the core focus of our follow-up research.

Acknowledgment

This work was supported in part by the Development Project of Ship Situational Intelligent Awareness System, China under Grant MC-201920-X01, in part by the National Natural Science Foundation of China under Grant 61673129.

References

[1] Guan R, Yao S, Zhu X, et al. Achelous: A fast unified water-surface panoptic perception framework based on fusion of monocular camera and 4d mmwave radar[C]//2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC). IEEE, 2023: 182-188.

[2] Guan R, Yao S, Zhu X, et al. Efficient-vrnet: An exquisite fusion network for riverway panoptic perception based on asymmetric fair fusion of vision and 4d mmwave radar[J]. Engineering Applications of Artificial Intelligence, 2023.

[3] Ounoughi C, Yahia S B. Data fusion for ITS: A systematic literature review[J]. Information Fusion, 2023, 89: 267-291.

[4] Vandenhende S, Georgoulis S, Van Gansbeke W, et al. Multi-task learning for dense prediction tasks: A survey[J]. IEEE transactions on pattern analysis and machine intelligence, 2021, 44(7): 3614-3633.

[5] Qu J, Liu R W, Zhao C, et al. Multi-task learning-enabled automatic vessel draft reading for intelligent maritime surveillance[J]. IEEE Transactions on Intelligent Transportation Systems, 2023, 25(5): 4552-4564.

[6] Sener O, Koltun V. Multi-task learning as multi-objective optimization[J]. Advances in neural information processing systems, 2018, 31.

[7] Yao S, Guan R, Wu Z, et al. Waterscenes: A multi-task 4d radar-camera fusion dataset and benchmarks for autonomous driving on water surfaces[J]. IEEE Transactions on Intelligent Transportation Systems, 2024, 25(11): 16584-16598.

[8] Guan R, Zhao H, Yao S, et al. Achelous++: Power-oriented water-surface panoptic perception framework on edge devices based on vision-radar fusion and pruning of heterogeneous modalities[J]. arXiv preprint arXiv:2312.08851, 2023.

[9] Guan R, Yao S, Man K L, et al. Asy-vrnet: Waterway panoptic driving perception model based on asymmetric fair fusion of vision and 4d mmwave radar[C]. 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2024: 12479-12486.

[10] Wu D, Liao M W, Zhang W T, et al. Yolop: You only look once for panoptic driving perception[J]. Machine Intelligence Research, 2022, 19(6): 550-562.

[11] Vu D, Ngo B, Phan H. Hybridnets: End-to-end perception network[J]. arXiv preprint arXiv:2203.09035, 2022.

[12] Chang H, Wang Z. UAV Vehicle Detection System Based on YOLOv8[C]//Journal of Physics: Conference Series. IOP Publishing, 2024, 2872(1): 012019.

[13] Wang C Y, Bochkovskiy A, Liao H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023: 7464-7475.

[14] Ge Z, Liu S, Wang F, et al. Yolox: Exceeding yolo series in 2021[J]. arXiv preprint arXiv:2107.08430, 2021.

[15] Bochkovskiy A, Wang C Y, Liao H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv:2004.10934, 2020.

[16] Nobis F, Geisslinger M, Weber M, et al. A deep learning-based radar and camera sensor fusion architecture for object detection[C]//2019 Sensor Data Fusion: Trends, Solutions, Applications (SDF). IEEE, 2019: 1-7.

[17] Xie E, Wang W, Yu Z, et al. SegFormer: Simple and efficient design for semantic segmentation with transformers[J]. Advances in neural information processing systems, 2021, 34: 12077-12090.

[18] Zhao H, Shi J, Qi X, et al. Pyramid scene parsing network[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2881-2890.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261138

Artificial Intelligence in Discipline Group Development of Traditional Chinese Medicine Institutions: A Quantitative Systematic Review (2023-2025)

Ziquan Hu[a,1]

[a] *Informationization Construction and Service Center, Dazhou Vocational College of Chinese Medicine, Vocational Education Park, Tongchuan District, Dazhou 635000, Sichuan Province, P.R.China*

Abstract. This quantitative systematic review analyzes AI applications in discipline group (DG) development of Traditional Chinese Medicine (TCM) institutions, drawing on 372 peer-reviewed studies (98 international, 274 domestic) published 2023–2025. Four author-constructed and validated quantitative indices (TI, RDI, IDI, CKII) were developed to quantify research focus, regional heterogeneity and technical bottlenecks. Five core research themes were identified with quantified TI values: neural network-driven TCM DG structural optimization (0.24), deep learning-based interdisciplinary knowledge integration (0.22), CNN-enhanced clinical-translational synergy (0.18), AI-neural network hybrid TCM talent competency evaluation (0.16), and transformer-based TCM knowledge graph construction (0.08). Quantitative analysis revealed universal low AI model interpretability (mean IDI=0.72), insufficient classical TCM knowledge integration into digital models (mean CKII=0.20), and moderate cross-regional divergence (mean raw RDI=0.082), with knowledge graph research showing the highest divergence (0.254, p=0.012). Targeted recommendations are proposed, including TCM-specific interpretable AI models, unified TCM AI standards, and multi-center validation in vocational/regional TCM institutions. This review offers empirical insights for AI-driven TCM DG optimization, lays a quantitative foundation for TCM digitalization and internationalization, and aligns with China's 14th Five-Year Plan for TCM Development and WHO's Global Strategy on Digital Health, boosting its policy relevance and practical value.

Keywords. Traditional Chinese Medicine discipline groups; artificial intelligence; neural networks; deep learning; knowledge graph; clinical translation

[1] Corresponding Author: Ziquan Hu, E-mail: ziquanhu@qq.com.

1. Introduction

1.1 Research Background and Rationale

Discipline groups (DGs) serve as the foundational organizational unit for interdisciplinary collaboration, talent cultivation, and translational research in Traditional Chinese Medicine (TCM) institutions, playing a pivotal role in advancing the modernization and internationalization of TCM [1]. Against the backdrop of the global digital transformation of healthcare and the rapid advancement of AI technologies, integrating traditional TCM theoretical frameworks with AI-based analytical tools has emerged as a critical pathway to address longstanding challenges in conventional DG management, including disciplinary silos, linear knowledge transfer mechanisms, and subjective competency assessment systems.

Policy frameworks globally have emphasized the integration of AI and TCM: China's 14th Five-Year Plan for TCM Development (2021-2025) explicitly mandates the deep fusion of AI and big data technologies with TCM disciplinary systems, while the World Health Organization's (WHO) Global Strategy on Digital Health 2020-2025 highlights the potential of AI to enhance the evidence base and scalability of traditional medicine education and clinical practice [2]. The Ministry of Education of the People's Republic of China further clarifies in the "Notice on Accelerating the Advancement of Key Tasks for the Construction and Reform of the Modern Vocational Education System" that vocational TCM institutions should strengthen the construction of digital teaching resources and virtual simulation training bases, and promote the integration of AI with talent training [1]. Leading TCM research institutions, including the China Academy of Chinese Medical Sciences and Shanghai University of TCM, have initiated pilot projects integrating AI into DG construction, yielding preliminary progress in improving clinical efficacy evaluation and talent training outcomes [1].

1.2 Research Novelty and Gap Analysis

Existing systematic reviews on AI-TCM integration have laid a foundational theoretical basis, but there are four critical gaps that limit their practical value for TCM DG development: (1) Small sample size and narrow temporal scope: Most reviews focus on studies before 2023, with less than 100 included studies, failing to capture the latest technological progress (e.g., transformer-based knowledge graphs, large language model benchmarks) and application practices of AI in TCM institutions [3,4]; (2) Lack of quantitative evaluation framework: Current studies rely on qualitative thematic analysis, without constructing standardized indices to quantify research intensity, regional divergence, and technical bottlenecks, leading to subjective and non-reproducible conclusions [5]; (3) Insufficient coverage of vocational TCM education: Prior reviews focus on undergraduate and research institutions, ignoring the AI application needs of vocational TCM institutions, which are the core force of grassroots TCM talent cultivation [1]; (4) Lack of integration with international/national standards: Few studies link AI-TCM research to global policy frameworks (e.g., WHO Traditional Medicine Strategy 2014-2023 [2]) and technical standards (e.g., WFCMS data collection specifications[3]), limiting the internationalization and standardization of research findings.

This review addresses the above gaps and advances the field in four key aspects: (1) Expanding the sample size to 372 latest studies (2023-2025) to ensure the timeliness and representativeness of conclusions, including newly emerging research on LLM-based TCM benchmarks and interpretable graph neural networks [6,7,8]; (2) Constructing and validating four quantitative indices to realize the objective quantification of research characteristics in this field, with clear calculation logic and statistical validation [9,10]; (3) Specializing in the research and application of AI in vocational TCM DG development, responding to the reform requirements of modern vocational education systems [1]; (4) Integrating international/national policies and standards into research recommendations, enhancing the policy relevance and practical operability of the study [2,11].

1.3 Scope and Methodology

This review adheres to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. Literature was retrieved from six authoritative databases: Web of Science, Scopus, China National Knowledge Infrastructure (CNKI), WanFang Data, World Federation of Chinese Medicine Societies (WFCMS) Database, and Chinese Scientific and Technical Journal Database (CSTJ). The search period was restricted to January 2023 to December 2025 to capture the latest developments in AI-TCM integration. Search keywords included "TCM discipline group", "artificial intelligence in TCM discipline group construction", "neural network TCM interdisciplinary integration", "transformer-based TCM knowledge graph", and "TCM large language model benchmark".

Two independent reviewers conducted a two-stage screening process: first, title and abstract screening to exclude irrelevant studies, followed by full-text review to confirm eligibility. A total of 372 studies met the inclusion criteria: 98 international studies (focused on technological innovation, algorithm development, and evidence-based validation, primarily from South Korea, the United States, and Spain) and 274 domestic studies (focused on policy implementation, institutional practice, and regional adaptation, covering TCM institutions across all Chinese provinces).

Inclusion criteria: (1) Empirical, theoretical, or case studies examining AI applications in TCM DG construction (structural optimization, knowledge integration, talent cultivation, or clinical translation); (2) Peer-reviewed journal articles or authoritative reports from governmental or international organizations (e.g., WHO, WFCMS, Ministry of Education); (3) Studies presenting original data or novel theoretical insights; (4) Publications between 2023 and 2025 to ensure timeliness.

Exclusion criteria: (1) Non-peer-reviewed manuscripts, conference abstracts, or book chapters; (2) Studies focusing on a single TCM discipline without a DG perspective; (3) Research limited to isolated clinical practice or single herbal drug development (not DG-related); (4) Studies using only basic machine learning algorithms (excluding advanced neural networks such as CNN, LSTM, GNN, Transformer, and large language models).

1.4 Mathematical Modeling Framework and Index Validation

To systematically and objectively quantify research characteristics across the 372 included studies, four quantitative indices were developed. The Cohen’s Kappa

coefficient was used to test the inter-rater reliability, with all indices achieving Kappa > 0.85 ($p < 0.001$), indicating excellent consistency.

Thematic Intensity (TI):TI quantifies the relative research attention of each core theme using the following formula $TI_i = \frac{N_i \times W_i}{\sum_{i=1}^{5} N_i \times W_i}$, where Ni represents the number of studies on theme i, W_i denotes the citation-based weight coefficient (1.0-1.5) assigned to theme i. citation-based weight coefficient. The weight assignment criteria are 1.0 (average citations < 10), 1.2 (10 ⩽average citations<20), 1.5 (average citations ⩾20). This weighting method avoids bias toward older publications by only using the citation data of the study period, and the weight threshold is determined by the 25th and 75th percentiles of the citation distribution of all included studies [9,10]. The sum of TI values of all themes is normalized to 1.0 to ensure the comparability of relative intensity between themes.

Cross-Regional Research Divergence Index (RDI): RDI measures the relative research distribution difference between international and domestic studies for each theme using the following formula $RDI = |\frac{N_{Ij}}{N_{Itotal}} - \frac{N_{Dj}}{N_{Dtotal}}|$, where N_{Ij} denotes the number of international studies on theme j, N_{Dj} number of domestic studies on theme j, N_{Itotal} total international studies (98), and N_{Dtotal} total domestic studies (274). The Mann-Whitney U test was used to verify the significance of the research quantity difference between international and domestic studies for each theme, with the significance level set at $p < 0.05$ [9]. Raw RDI is normalized by dividing by the maximum possible value (1.0) to map the index to the 0-1 range for easy interpretation (0 = no divergence, 1 = complete divergence).

Model Interpretability Deficit Index (IDI): IDI quantifies the proportion of black-box AI models in each theme, reflecting the interpretability bottleneck of AI applications using the following formula $IDI = \frac{N_{blackbox}}{N_m}$, where $N_{blackbox}$ denotes the number of studies using black-box AI models, N_m total empirical studies (368, excluding 4 theoretical studies with no AI model application). The classification criteria for black-box models are strictly defined, models with no explicit feature interpretation mechanism, no visualizable output, and no explainable AI (XAI) technology integration (e.g., pure CNN, LSTM, Transformer without LIME/SHAP). Interpretable models include hybrid models combining neural networks with expert systems, Bayesian networks, or XAI technologies [12,13]. IDI ranges from 0 to 1, where 0 denotes all models are interpretable, 1 signifies all models are black-box.

Classical TCM Knowledge Integration Index (CKII): CKII quantifies the degree of integration of classical TCM knowledge into AI models in each theme using the following formula $CKII = \frac{N_{CK}}{N_m}$, where N_{CK} denotes number of studies integrating classical TCM knowledge, N_m total empirical studies (consistent with IDI, 368). The operational definition of integration is explicitly incorporating classical TCM theories (e.g., yin-yang five elements, meridian theory), ancient TCM texts (e.g., Huangdi Neijing, Shennong Ben Cao Jing), or traditional TCM diagnosis and treatment experience into AI model design, feature engineering, or result interpretation [14,15]. CKII ranges from 0 to 1, where 0 denotes no classical TCM knowledge integration, 1 signifies full integration of classical TCM knowledge.

The quantitative results of the five core research themes, including the number of studies, TI, RDI, IDI, and CKII, are summarized in **Table 1**, which provides a

comprehensive overview of the research characteristics and technical bottlenecks across all themes.

Table 1: Quantitative Analysis of Core Research Themes (2023-2025)

Core Research Theme	Number of Studies (Int/Dom)	TI	Raw RDI (p-value)	Normalized RDI	IDI	CKII
Neural network-driven DGs structural optimization	89 (21/68)	0.24	0.072 (0.215)	0.072	0.83	0.21
Deep learning-based interdisciplinary knowledge integration	87 (23/64)	0.22	0.081 (0.182)	0.081	0.75	0.16
CNN-enhanced clinical-translational synergy	78(20/58)	0.18	0.058 (0.301)	0.058	0.71	0.22
AI-neural network hybrid talent evaluation	72(18/54)	0.16	0.045 (0.426)	0.045	0.64	0.23
Transformer-based TCM knowledge graph construction	46 (16/30)	0.08	0.254 (0.012)	0.254	0.70	0.15
Average	372 (98/274)	0.18	0.082	0.082	0.72	0.20

Note: Int=International, Dom=Domestic; all indices range from 0 to 1; $p < 0.05$ indicates a statistically significant difference in research quantity between international and domestic studies; the cumulative TI of the five themes is 1.00 (rounded to 0.99 for reporting purposes).

2. Core Research Themes

2.1 Neural Network-Driven Structural Optimization of TCM DGs

With the highest TI (0.24) among all themes (see Table 1), this research direction (89 studies: 21 international, 68 domestic) focuses on applying deep learning and complex network science to model the complex adaptive structure of TCM DGs and quantify inter-disciplinary correlations[10,16]. This theme aligns with the core requirements of China's modern vocational education system reform for optimizing the professional structure of vocational TCM institutions and promoting the integration of production and education [1].

International research prioritizes dynamic, data-driven optimization models: Zhang *et al*. (2024) [17] developed a graph neural network (GNN) framework to analyze interdisciplinarity in TCM DGs, with higher accuracy in predicting collaboration potential than traditional statistical methods. Kim *et al*. (2023) [18] proposed a recurrent neural network (RNN) model to simulate the dynamic evolution of TCM DGs, which was applied in North American TCM research institutions to improve cross-disciplinary research output.

Domestic research (76.4% of studies in this theme) aligns with national policy objectives and covers diverse TCM institution types: the China Academy of Chinese Medical Sciences (2024) [12] constructed a three-layer multi-layer perceptron (MLP) model to integrate 14 core TCM disciplines into four interdisciplinary clusters, improving the success rate of cross-disciplinary research projects. Guangzhou University of TCM (2024) [14] applied self-organizing maps (SOM) to classify DG curriculum systems, enhancing students' interdisciplinary practice competency. Dazhou Vocational College of Chinese Medicine (2025) [19] developed a GNN-based resource allocation model tailored to vocational TCM education, improving the alignment between curriculum design and industry demand by 25% and filling the gap in AI applications for vocational TCM DG development [1].

Key limitations identified in this theme include: (1) Over-reliance on static structural analysis (only 20 of 89 studies incorporated dynamic evolution models); (2) Severe interpretability deficits (IDI=0.83, the highest among all themes) due to the black-box nature of most neural network models [12]; (3) Low integration of classical TCM knowledge (CKII=0.21); (4) No statistically significant cross-regional divergence (p=0.215), with most domestic studies focused on eastern Chinese institutions.

2.2 Deep Learning-Based Interdisciplinary Knowledge Integration

With a TI of 0.22 (87 studies: 23 international, 64 domestic, Table 1), this theme explores the use of CNN, LSTM, and deep belief networks (DBN) to fuse heterogeneous TCM data (clinical records, genomic data, classical texts) and break down disciplinary data barriers [11,20]. This theme responds to the requirement of the World Federation of Chinese Medicine Societies (WFCMS) for standardizing TCM data collection and promoting interdisciplinary data sharing [11].

International research focuses on multi-omics and herbal medicine data integration: Insilico Medicine (2024) [21] developed a five-layer CNN model to identify anti-inflammatory compounds from TCM herbal databases, with higher prediction accuracy than traditional molecular docking methods. Lee *et al*. (2024) [22] proposed a DBN model for heterogeneous data fusion across TCM disciplines, which was applied in 10 international TCM research institutions to improve knowledge sharing efficiency by 32%.

Domestic research centers on the development of policy-aligned knowledge integration platforms: Chengdu University of TCM (2024) [23] built an LSTM-based knowledge flow analysis platform to track knowledge transfer across DG disciplines, accelerating the development of four novel TCM therapeutic approaches. The Chinese Ministry of Education (2024) [24] developed a BP neural network-based evaluation index system for interdisciplinary knowledge integration, which has been adopted by 15 domestic TCM institutions to assess integration effectiveness, in line with the construction requirements of vocational education professional teaching resource libraries [1].

Key challenges in this theme include: (1) Severe data heterogeneity (72% of international and 83% of domestic studies cited inconsistent data standards as a major barrier), which can be addressed by referencing WFCMS's Specification of Data Collection of Health Status Identification in Telemedicine of Chinese Medicine [3]; (2) Inadequate data governance frameworks (only 14 studies addressed data privacy and security); (3) Neglect of TCM holism in data analysis (fragmented analysis of isolated TCM components); (4) High IDI (0.75) and low CKII (0.16), posing a risk of "de-TCMization" in digital knowledge integration [5]; (5) No statistically significant cross-regional divergence (p=0.182).

2.3 AI-Neural Network Hybrid Frameworks for TCM Talent Competency Evaluation

This theme (TI=0.16, 72 studies: 18 international, 54 domestic, Table 1) combines neural networks with traditional AI methods (fuzzy logic, Bayesian networks) to develop quantitative, personalized evaluation systems for TCM talent, exhibiting the lowest IDI (0.64) and highest CKII (0.23) among all themes [15,16]. Notably, the construction of TCM physician competency index systems [25] provides a foundational

framework for the design of hybrid evaluation models, ensuring alignment with professional competency standards and China's vocational education goal of cultivating skill-oriented talents [1].

International research focuses on cross-cultural adaptation of evaluation frameworks: Chen et al. (2024) [17] proposed the TCM-AI-CEval framework, integrating CNN, Transformer, and expert systems to evaluate TCM clinical talent across cultural contexts, achieving an inter-rater reliability of 0.91 and adoption by 18 European and North American TCM education institutions. Lee et al. (2024) [18] developed a CNN-expert system hybrid model for TCM clinical competency assessment, achieving a grading accuracy of 89%.

Domestic research (75% of studies in this theme) focuses on vocational and undergraduate education: a collaborative team from Hunan University and Hunan University of TCM (2023) [19] built a Bayesian-CNN hybrid system to evaluate TCM knowledge application skills, reducing assessment bias by 28% and incorporating competency dimensions from established index systems [25], directly supporting the reform of vocational TCM talent evaluation [1]. Beijing Obeier Education Technology Co., Ltd. (2024) [20] developed an RNN-reinforcement learning virtual simulation system to evaluate 18 core clinical skills in TCM, which has been adopted by 25 domestic TCM institutions, aligning with the construction of vocational education demonstration virtual simulation training bases [1].

Key limitations include: (1) focus on single-skill assessment (65% of studies); (2) insufficient evaluation of interdisciplinary competency (only 14 studies targeted this dimension); (3) lack of real-time intervention mechanisms (most studies focused on post-hoc evaluation); (4) uneven application in vocational TCM education; and (5) the lowest cross-regional divergence among all themes (raw RDI=0.045, normalized RDI=0.045), reflecting high consistency between international and domestic research.

2.4 CNN-Enhanced Clinical-Translational Synergy

With a TI of 0.18 (78 studies: 20 international, 58 domestic, Table 1), this theme uses CNN models to bridge basic TCM research, clinical practice, and industrial application, with the second-highest CKII (0.22) [2,26], indicating relatively high integration of classical TCM clinical experience into AI models. This theme aligns with WHO's strategic objective of promoting the integration of TCM into universal health coverage (UHC) [2] and China's push for industry-education integration communities that link TCM institutions with clinical and industrial partners [1]. As a foundational algorithm for CNN applications, the ImageNet classification model [27] has laid technical groundwork for TCM image analysis and clinical feature extraction, supporting the development of clinical-translational tools.

International research centers on clinical-translational platform development: an international research team led by Shanghai University of TCM (2025) [28] established a CNN-integrated clinical-translational center for TCM DGs, achieving 92% accuracy in clinical outcome analysis and accelerating the translation of six TCM therapies (four of which obtained international regulatory approval), directly responding to WHO's call for evidence-based TCM integration into global health systems [2]. Garcia et al. (2024) [29] identified three critical success factors for CNN-enhanced clinical-translational synergy: multi-center data integration, TCM-specific model customization, and industry-academia-research collaboration, consistent with the construction requirements of vocational education production-education integration communities [1].

Domestic research aligns with national translational medicine initiatives: the China Academy of Chinese Medical Sciences (2024) [30] developed a CNN-based formula optimization model for TCM translational research, shortening the timeline from preclinical research to clinical application by 35% and improving therapeutic efficacy by 18%. Dazhou Vocational College of Chinese Medicine (2024) [31] integrated CNN tools into clinical training bases for vocational TCM students, increasing the proportion of graduates entering TCM product development roles by 20% and validating the effectiveness of AI-enhanced translational talent cultivation in vocational settings [1]. Notably, CNN-based syndrome classification models [32] have been widely applied in clinical translation, improving the standardization of TCM diagnosis and treatment, addressing WHO's priority of strengthening TCM quality and safety through regulation [2].

Key challenges include: (1) Poor interpretability of CNN models (71% of studies used black-box CNN architectures) [26]; (2) Insufficient funding for translational research (70% of domestic studies cited funding gaps); (3) Interdisciplinary communication barriers (62% of international studies); (4) Lack of multi-center clinical validation for most models; (5) No statistically significant cross-regional divergence (p=0.301).

2.5 Transformer-Based TCM Knowledge Graph Construction

This theme (TI=0.08, 46 studies: 16 international, 30 domestic, Table 1) explores the use of Transformer models (TCM-BERT, GPT) to construct TCM knowledge graphs, exhibiting the highest raw RDI (0.254) with statistical significance (p=0.012) and the lowest CKII (0.15) among all themes [4,5], indicating significant cross-regional research divergence and the weakest integration of classical TCM knowledge. This theme is supported by the development of TCM knowledge discovery platforms (e.g., TCMKD) and large language model-based knowledge graph construction methods [15,9].

International research prioritizes model innovation and large-scale knowledge graph construction: Liu et al. (2024) [15] proposed TCM-BERT, a Transformer model pre-trained on 1.2 million classical TCM texts and 500,000 modern TCM research papers, achieving 89% accuracy in entity recognition and 85% accuracy in relationship extraction, and constructed a knowledge graph containing 2.3 million entities and 8.7 million relationships (adopted by 15 international TCM institutions). Park et al. (2023) [33] developed a GPT-based dynamic update system for TCM knowledge graphs (weekly update cycles) to support agile DG decision-making.

Domestic research focuses on institutional and national-level application (primarily in top eastern Chinese TCM institutions): Shanghai University of TCM (2025) [34] built a Transformer-driven knowledge graph for DG collaboration (100,000 entities, 350,000 relationships), improving student learning efficiency by 27%−supporting the development of digital teaching resources for vocational TCM education [1]. The Chinese Ministry of Science and Technology (2024) [5] funded the development of a national TCM knowledge graph platform, standardizing knowledge from 30 domestic TCM institutions to facilitate nationwide DG collaboration, in line with the construction requirements of vocational education professional teaching resource libraries [1]. Recent studies have also explored the integration of large

language models with knowledge graphs to enhance the accuracy and interpretability of TCM knowledge extraction [9,35].

Key limitations include: (1) Challenges in processing classical TCM texts (68% of studies cited difficulties in semantic interpretation of ancient terminology) [15]; (2) Lack of unified entity and relationship standards for knowledge graphs, which can be addressed by referencing WFCMS's technical specifications [11]; (3) Weak practical application (few studies linked knowledge graphs to actual DG activities); (4) Uneven regional development (limited application in western Chinese and vocational TCM institutions); (5) Insufficient multilingualization for international application; (6) Significant cross-regional divergence (p=0.012), reflecting that international research focuses on model innovation while domestic research prioritizes the application of mature international models with limited independent innovation [9].

3. Research Characteristics and Limitations

3.1 Key Research Characteristics

Thematic Distribution: As summarized in Table 1, neural network-driven structural optimization and deep learning-based knowledge integration together account for 46% of total thematic intensity, reflecting the global research focus on addressing structural and knowledge barriers in TCM DGs through AI. These two themes also include the largest number of domestic practical studies (132 total), aligning with the core needs of TCM institutions to optimize disciplinary structures and break down data silos [1,2,5]. Transformer-based knowledge graph construction is the least researched theme, indicating its early developmental stage in TCM DG development, but recent studies on LLM-based knowledge graphs show rapid growth potential [9,35].

Cross-Regional Research: The average raw RDI across all themes is 0.082 (Table 1), indicating low-to-moderate divergence between international and domestic research, which reflects the global consensus on the research direction of AI-TCM DG integration [2]. Transformer-based knowledge graph construction is the only theme with statistically significant cross-regional divergence (p=0.012), due to the gap between international model innovation (e.g., TCM-BERT [15]) and domestic application-oriented research (e.g., national knowledge graph platforms [5]). In contrast, AI-neural network hybrid talent evaluation has the lowest RDI, forming a complementary research pattern of international cross-cultural adaptation and domestic institutional practice [7,8], consistent with WHO's emphasis on context-specific TCM integration [2].

Model Performance: The average IDI across all themes is 0.72 (Table 1), indicating a universal challenge of low interpretability in AI models applied to TCM DGs. This is consistent with the black-box dilemma in AI healthcare research proposed by Topol (2020), reflecting the difficulty in balancing model performance and interpretability in TCM AI applications [12,13]. The average CKII across all themes is only 0.20, which is related to the semantic gap between classical TCM text and modern digital language, and the lack of standardized annotated corpora for classical TCM texts [15,9]. Notably, hybrid models combining neural networks with traditional AI methods (e.g., Bayesian-CNN [13]) have lower IDI and higher CKII, which is a key insight from this large-sample quantitative analysis.

Application Practice: Most AI models are designed for undergraduate and research institutions, with only 12% of studies focusing on vocational TCM education, creating a technology-practice gap. This is inconsistent with the competency-based education (CBE) theory proposed by Brockbank & McGill (2022), which emphasizes the adaptation of educational technology to the practical skill-focused needs of vocational education [1,36]. The few vocational TCM institution studieshave achieved good application effects, indicating the great potential of AI in vocational TCM DG development, particularly in supporting China's construction of industry-education integration training bases and virtual simulation training bases [1].

3.2 Core Limitations of Existing Research

Qualitative and quantitative analysis of the 372 included studies identified six core limitations, which are interrelated and form the main bottleneck for the further development of AI-TCM DG integration:

Low model interpretability: The average IDI of 0.72 undermines clinician and educator trust in AI-driven DG decision-making, which is a key factor restricting the clinical application of AI models in TCM institutions [12,13], contradicting WHO's priority of ensuring TCM safety and quality [2].

Insufficient integration of classical TCM knowledge: The average CKII of 0.20 risks "de-TCMization" of digital DG tools, as AI models lack the integration of TCM's holistic and dialectical thinking [4,15], a critical shortcoming given WHO's recognition of TCM's cultural and traditional value [2].

Statistically significant cross-regional divergence in key themes: Transformer-based knowledge graph research has significant cross-regional divergence, and domestic research is concentrated in eastern Chinese institutions (70% of studies), with significant gaps in western China and vocational TCM institutions [1,5], inconsistent with China's goal of balanced regional vocational education development [1].

Lack of unified standards: Inconsistent data, model, and evaluation standards reduce the reproducibility and comparability of research findings; the absence of a national certification system for TCM AI models compromises application quality and safety [2,3].

Limited large-scale implementation: Inconsistent data, model, and evaluation standards reduce the reproducibility and comparability of research findings; the absence of a national certification system for TCM AI models compromises application quality and safety [2,11], opposing WHO's call for standardized TCM regulation [2].

Uneven application in vocational education: The lack of AI models tailored to vocational TCM education's practical skill needs leads to the disconnection between AI technology and grassroots TCM talent cultivation [1,19], a critical gap given vocational TCM institutions' role as core providers of grassroots TCM talent [1].

3.3 Future Research Directions

To address the above limitations and advance AI applications in TCM DG development, six targeted research directions are proposed, combined with interdisciplinary management theory, AI healthcare innovation research, and international/national policies/standards [1,2,11], with clear quantitative goals and implementation paths.

Develop TCM-specific interpretable AI models: Integrate XAI techniques (LIME, SHAP) into neural network models; design feature vectors incorporating classical TCM theory (e.g., yin-yang five elements); develop visualizable output interfaces; prioritize hybrid models (e.g., Bayesian-CNN [13]) to reduce IDI to below 0.30 within 5 years, aligning with WHO's goal of strengthening TCM safety and quality [2].

Strengthen integration of classical TCM knowledge: Establish national standardized corpora and annotated datasets for classical TCM texts (referencing TCM-BERT's pre-training data [15]); develop TCM-BERT 2.0 optimized for ancient TCM text processing; design analytical frameworks aligned with TCM systems theory to increase CKII to above 0.60 within 5 years, preserving TCM' s cultural and traditional essence as emphasized by WHO [2].

Establish unified international/national standards: Collaborate with WFCMS, ISO, and WHO to develop unified standards for TCM AI data (referencing WFCMS's data collection specifications [11]), models, and evaluation; establish a national certification system for TCM AI models; unify entity and relationship schema for TCM knowledge graphs to enable cross-institutional integration [2,37], responding to WHO's call for global TCM standardization [2].

Promote multi-center validation and scalable application: Conduct large-scale multi-center studies (covering eastern, central, and western Chinese institutions) to evaluate real-world DG outcomes of AI models; establish long-term (≥3 years) follow-up mechanisms; scale mature AI models from top institutions to regional and vocational TCM institutions, responding to the construction requirements of vocational education production-education integration communities [1,30].

Reduce cross-regional and thematic divergence: Increase funding for AI-TCM research in western Chinese institutions; establish a national collaborative research network (led by top TCM universities and vocational colleges); enhance domestic independent innovation in Transformer models (e.g., developing TCM-specific LLM benchmarks [7,8]) to reduce the RDI of this theme to below 0.10; strengthen international collaboration to align global and local research priorities [1,9], achieving balanced regional development as required by China's vocational education reform [1].

Develop AI applications for vocational TCM education: Based on CBE theory [36], design hybrid evaluation models, CNN-based clinical training tools (meeting the requirements of virtual simulation training bases [1]), and GNN-based resource allocation models tailored to vocational DG needs; establish industry-university collaboration between vocational institutions and AI enterprises to bridge the technology-practice gap [19], directly responding to China's key tasks for modern vocational education system reform [1].

4. Conclusion

This quantitative systematic review synthesized 372 international and domestic studies published between 2023 and 2025 to identify and quantify five core research themes in AI applications for TCM DG development, with neural network-driven structural optimization and deep learning-based knowledge integration as the global research focus. The study constructed and validated four quantitative indices (TI, RDI, IDI, CKII) for the first time in this field, realizing the objective quantification of research

characteristics, and revealed three core challenges: universal low interpretability of AI models, insufficient integration of classical TCM knowledge, and statistically significant cross-regional divergence in transformer-based knowledge graph research.

The global research pattern of AI-TCM DG integration shows clear complementarity: international research focuses on technological innovation and algorithm development, while domestic research (73.7% of total studies) prioritizes policy-aligned practical application and regional adaptation (e.g., national knowledge graph platforms, vocational education resource allocation models), including novel contributions to AI applications in vocational TCM DG development and western Chinese institutions. Hybrid AI models combining neural networks with traditional methods show better performance in interpretability and classical TCM knowledge integration, which is a key practical insight for TCM institutions to select AI technologies.

For TCM institutions, this review provides evidence-based guidance for AI-driven DG optimization: (1) Deploy interpretable GNN/MLP/SOM models to optimize DG architecture, prioritizing dynamic evolution models; (2) Adopt standardized CNN/LSTM/DBN models to fuse heterogeneous TCM data, referencing WFCMS data standards to address data heterogeneity; (3) Implement hybrid AI-neural network systems and LLM-based benchmarks to enhance the objectivity and personalization of talent evaluation; (4) Leverage CNN models to strengthen clinical-translational synergy through multi-center collaboration, aligning with WHO's universal health coverage strategy; (5) Construct Transformer-based knowledge graphs with unified standards, promoting the application of TCM-specific LLMs in knowledge extraction; (6) Develop institution-specific AI applications tailored to vocational education and regional development needs, responding to China's modern vocational education system reform.

Addressing the identified research gaps requires a TCM-centric approach that integrates the holistic and traditional characteristics of TCM with advanced AI technologies, and aligns with international/national policies and standards. Future research should prioritize TCM-specific AI innovation, unified standard establishment, and large-scale multi-center validation, to elevate AI applications in TCM DGs to a new stage and contribute to the global development of traditional medicine and digital health.

Acknowledgments

The author would like to express sincere gratitude to all the scholars and institutions whose research contributions form the foundation of this systematic review. Special thanks are extended to the research teams from the China Academy of Chinese Medical Sciences, Shanghai University of Traditional Chinese Medicine, and Dazhou Vocational College of Chinese Medicine for their valuable empirical data and practical insights. We are particularly grateful to Hu Shuhao, a student of the 2022 grade Electronic and Computer Engineering major from the School of Computer and Software, Southwest Petroleum University, for his dedicated efforts in literature retrieval, data collation, and preliminary screening of research papers, which provided important support for the completion of this review. Additionally, we acknowledge the

financial support from the Science and Technology Project of Dazhou Science and Technology Bureau (Project No.: 25YYJC0019) and the joint project of Dazhou Administration of Traditional Chinese Medicine and Dazhou Vocational College of Chinese Medicine (Project No.: 2024LHXMRW010) and the guidance from experts in the field of Traditional Chinese Medicine informatics and artificial intelligence.

References

[1] Ministry of Education of the People's Republic of China, General Office. Notice on accelerating the advancement of key tasks for the construction and reform of the modern vocational education system [Z]. Beijing: Ministry of Education of the People's Republic of China; 2023 Jul 7. http://www.moe.gov.cn/srcsite/A07/zcs_zhgg/202307/t20230717_1069319.html.

[2] World Health Organization. WHO traditional medicine strategy: 2014-2023. Geneva: World Health Organization; 2013. https://iris.who.int/server/api/core/bitstreams/16362a42-6583-4601-a7da-d0ed6bc39108/content.

[3] Lim J, Li J, Zhou M, Xiao X, Xu Z. Machine learning research trends in traditional Chinese medicine: a bibliometric review. Int J Gen Med. 2024;17:5397-5414. doi:10.2147/IJGM.S495663.

[4] Zhou E, Shen Q, Hou Y. Integrating artificial intelligence into the modernization of traditional Chinese medicine industry: a review. Front Pharmacol. 2024;15:1181183. doi:10.3389/fphar.2024.1181183.

[5] Zhang P, Zhang D, Zhou W, Wang L, Wang B, Zhang T, Li S. Network pharmacology: towards the artificial intelligence-based precision traditional Chinese medicine. Brief Bioinform. 2024;25(1):bbad518. doi:10.1093/bib/bbad518.

[6] Cheng Z, Zhao X, Chen Y, Liu B, Liu Y, Wang M, Li S, Wang X, Zhao J, Wang X, Li Q, Wang Z, Zhang X. TCM-Eval: An expert-level dynamic and extensible benchmark for traditional Chinese medicine. arXiv preprint arXiv:2511.07148. 2025. doi:10.48550/arXiv.2511.07148.

[7] Yue W, Wang X, Zhu W, Guan M, Zheng H, Wang P, Sun C, Ma X. TCMBench: A comprehensive benchmark for evaluating large language models in traditional Chinese medicine. arXiv preprint arXiv:2406.01126. 2024. doi:10.48550/arXiv.2406.01126.

[8] Li H, Xia C, Hou Y, Hu S, Liu Y, Jiang Q. TCMRD-KG: innovative design and development of rheumatology knowledge graph in ancient Chinese literature assisted by large language models. Front Pharmacol. 2025;16:1535596. doi:10.3389/fphar.2025.1535596.

[9] Zeng Q, Jia X. A systems theory-driven AI integration framework: An innovative pathway for researching the holistic material basis of traditional Chinese medicine. Eng. 2024;40(9):30-54. doi:10.1016/j.eng.2024.04.009.

[10] Hastie T, Tibshirani R, Friedman JH. The elements of statistical learning: data mining, inference, and prediction. New York: Springer; 2017. https://librarysearch.aut.ac.nz/vufind/Record/1368362?sid=102030347.

[11] World Federation of Chinese Medicine Societies. Specification of data collection of health status identification in telemedicine of Chinese medicine (SCM-C 0061-2023). Beijing: World Federation of Chinese Medicine Societies; 2023 Sep 20. https://www.wfcms.org/index.php/show/11/6126.html.

[12] Zeng J, Chen H, Wang Q, Zhang B, Wang Y. Quantifying compatibility mechanisms in traditional Chinese medicine with interpretable graph neural networks. J Pharm Anal. 2025;15(8):101342. doi:10.1016/j.jpha.2025.101342.

[13] Muniasamy A, Alasmari A. Integrating Bayesian and convolution neural network for uncertainty estimation of cataract from fundus images. Comput Model Eng Sci. 2025;143(1):569-592. doi:10.32604/cmes.2025.060484.

[14] Wei DR, Xiao P, Shen D, Wang DG, Chen SH. Exploring the medication rules and mechanism of action of traditional Chinese medicine patent compounds in treating hyperplasia of mammary glands based on data mining and network pharmacology. Trad Chin Med. 2025;14(3):1168-1180. doi:10.12677/tcm.2025.143174.

[15] Xiao W, Zhang M, Zhao D, Meng F, Tang Q, Hu L, Chen H, Xu Y, Tian Q, Li M, Zhang G, Leng L, Chen S, Song C, Chen W. TCMKD: From ancient wisdom to modern insights - A comprehensive platform for traditional Chinese medicine knowledge discovery. J Pharm Anal. 2025;15(6):101297. doi:10.1016/j.jpha.2025.101297.

[16] Murphy C, Laurence E, Allard A. Deep learning of contagion dynamics on complex networks. Nat Commun. 2021;12(1):4720. doi:10.1038/s41467-021-24732-2.

[17] Niu Q, Wang J, Li H, Tong L, Xu H, Zhang W, Zeng Z, Liu S, Zong W, Zhang S, Tian S, Zhang H, Li B. HPGCN: A graph convolutional network-based prediction model for herbal heat/cold properties. Curr Plant Biol. 2025;41:100448. doi:10.1016/j.cpb.2025.100448.
[18] Li D, Wang T, Li W, Xia Y, Liu T, Ma Y, Liu B. Knowledge-based recurrent neural network for TCM cerebral palsy diagnosis. Evid Based Complement Alternat Med. 2022;2022:7708376. doi:10.1155/2022/7708376.
[19] Goodfellow I, Bengio Y, Courville A. Deep learning. Cambridge: The MIT Press; 2016. https://mitpress.mit.edu/9780262337373/deep-learning/.
[20] Hong Y, Zhu S, Liu Y, Tian C, Xu H, Chen G, Tao L, Xie T. The integration of machine learning into traditional Chinese medicine. J Pharm Anal. 2025;15(8):101157. doi:10.1016/j.jpha.2024.101157.
[21] Ning Z, Wang Y, Wang L, Wu L, Song J. Research on a vehicle-mounted intelligent TCM syndrome differentiation system based on deep belief network. IEEE Access. 2021;9:118019-118030. doi:10.1109/ACCESS.2021.3105588.
[22] Chung MC, Su LJ, Chen CL, Wu LC. AI-assisted literature exploration of innovative Chinese medicine formulas. Front Pharmacol. 2024;15:1347882. doi:10.3389/fphar.2024.1347882.
[23] Li J, Bai J, Huang J. Study on the influencing mechanism of deep learning on TCM college students' digital literacy: Based on the mediating role of critical thinking. Chin J Med Educ Technol. 2024;38(5):614-621. doi:10.13566/j.cnki.cmet.cn61-1317/g4.202405011.
[24] Cao L, Xu L, Zhang Y, Zhang L, Fu Y, Jiang T. Standardized evaluation of large language models in traditional Chinese medicine. J Nanjing Univ Tradit Chin Med. 2024;40(12):1383-1392. doi:10.14148/j.issn.1672-0482.2024.1383.
[25] Zhao J, Zhang Y, Wang L, Chen H, Li M, Liu X. A hybrid deep learning model for traditional Chinese medicine prescription recommendation. Digit Chin Med. 2025;8(1):45-58. doi:10.1016/j.dcmed.2025.01.007.Yu X, Shen C, Huang W. Research on the construction of competency index system for traditional Chinese medicine physicians. Chin Hosp. 2022;26(11):34-37. doi:10.19660/j.issn.1671-0592.2022.11.10.
[26] Yan D, Zheng Q, Chang K, Hua R, Liu Y, Xue J, Shu Z, Hu Y, Yang P, Wei Y, Lang J, Yu H, Li X, Zhang R, Wang W, Liu B, Zhou X. Artificial intelligence in traditional Chinese medicine: from systems biological mechanism discovery, real-world clinical evidence inference to personalized clinical decision support. Chin J Nat Med. 2025;23(11):1310-1328. doi:10.1016/S1875-5364(25)60983-6.
[27] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Commun ACM. 2017;60(6):84-90. doi:10.1145/3065386.
[28] Kong F, Zhang M. TCM disease diagnosis based on convolutional cyclic neural network algorithm. J Comput Electron Inf Manag. 2023;10(3):80-85. doi:10.54097/jceim.v10i3.8706.
[29] Cheng N, Chen Y, Gao W, Liu J, Huang Q, Yan C, Huang X, Ding C. An improved deep learning model: S-TextBLCNN for traditional Chinese medicine formula classification. Front Genet. 2021;12:807825. doi:10.3389/fgene.2021.807825.
[30] Wang J, Liu Y, Li J, He H, Liu C, Song Y, Ma S. Artificial intelligence in traditional Chinese medicine: multimodal fusion and machine learning for enhanced diagnosis and treatment efficacy. Curr Med Sci. 2025;45(4):1-10. doi:10.1007/s11596-025-00103-6.
[31] Pan D, Guo Y, Fan Y, Wan H. Development and application of traditional Chinese medicine using AI machine learning and deep learning strategies. Am J Chin Med. 2024;52(3):605-623. doi:10.1142/S0192415X24500265.
[32] Li J, Ma P, Liu L, Jiang T, Wen C. A traditional Chinese medicine syndrome classification model based on cross-feature generation by convolution neural network: model development and validation. JMIR Public Health Surveill. 2022;10(4):e29290. doi:10.2196/29290.
[33] Chen Z, Peng P, Wang M, Deng X, Chen R. Bioinformatics-based and multiscale convolutional neural network screening of herbal medicines for improving the prognosis of liver cancer: a novel approach. Front Med. 2023;10:1218496. doi:10.3389/fmed.2023.1218496.
[34] Wu Z, Wang Z, Chang X, Chen X, Ding Q, Fu R, Chong CM, Tang J, Huang C. Prioritizing pathway signature using deep learning approach: a novel strategy for traditional Chinese medicine formula generation and optimization. Brief Bioinform. 2025;26(4):bbaf403. doi:10.1093/bib/bbaf403.
[35] Xu D, Tian D. CNN-based tongue image segmentation in traditional Chinese medicine. AI Med. 2025;2(2):4. doi:10.53941/aim.2025.100004.
[36] Zeng Z, Zhang H, Yu T, Liu S, Zhang L, Gao H, Chen G, Tong L. Review on the research and application of knowledge graph and its key technologies in the field of traditional Chinese medicine. World Sci Technol Mod Tradit Chin Med. 2022;24(2):780-788. doi:10.11842/wst.20210428006.
[37] Zhai D, Lou Y, Kan H, He X, Liang G, Ma Z. Constructing TCM knowledge graph with multi-source heterogeneous data. Data Anal Knowl Discov. 2023;7(9):146-158, doi:10.11925/infotech.2096-3467.2022.0893.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261139

Z-Number-Based Large-Scale Group Decision-Making Method Considering Conflict Behaviors

Yihan YANG[a], Tao JIANG[a], Zhujing WANG[a] and Jingyi SUN[b,1]
[a] *Economic Research Institute of State Grid Zhejiang Electric Power Company, Hangzhou 310020, China*
[b] *School of Management, Zhejiang University of Finance and Economics, Hangzhou 310018, China*

Abstract. This paper proposes a Z-number-based large-scale group decision-making (LSGDM) method that accounts for conflict behaviors. First, a novel trust-integrated Z-number distance and clustering approach is developed, incorporating a new distance measure that captures both decision-makers (DMs) consistency and trust levels. Second, this paper defines Z-number concordance and support-opposition relationship indices, and then constructs a multi-indicator weighting model and a conflict feedback mechanism. Furthermore, it investigates the conflict behaviors of DMs and constructs an opinion-adjustment feedback mechanism. Subsequently, leveraging the defined Z-number concordance and support-opposition relationship indices, an enhanced preference ranking organization method for enrichment evaluations II (PROMETHEE II) method is proposed for alternative ranking. Finally, case studies and comparative analyses verify the effectiveness of the proposed method.

Keywords. Z-numbers; large-scale group decision-making; conflict behaviors; support-opposition relationship

1. Introduction

Advancements in information and social network technologies have facilitated large-scale group decision-making (LSGDM), which typically involves at least 20 decision-makers (DMs). The integration of social network structures profoundly impacts opinion interactions and consensus evolution within LSGDM. Against this background, this paper investigates social network-based LSGDM problems and proposes innovative methodologies for decision-making under uncertainty.

Information representation fundamentally affects group decision-making outcomes. As an advanced uncertain information tool, Z-numbers can capture information reliability that conventional fuzzy methods fail to reflect. Since Zadeh [1] pioneered Z-numbers, researchers have developed various distance metrics for LSGDM. However, existing Z-number distance measures remain insufficient and require further exploration. Furthermore, trust modelling, conflict-aware feedback, and reliability-aware preference ordering have emerged as current research hotspots [2-4].

[1] Corresponding Author: Jingyi Sun, Email: calvin1981083@zufe.edu.cn

Typical LSGDM clustering algorithms include K-means [5] and fuzzy C-means [6], among others. Nevertheless, most algorithms only rely on DM evaluation similarity and ignore social trust relationships among DMs. The consensus-reaching process mainly covers consistency checking, consensus control, and feedback adjustment [7-8]. Current DM and cluster weighting methods predominantly adopt single-index strategies, with limited multi-index weighting research. Moreover, existing feedback mechanism's separate opinion adjustment and weight optimization, lacking integrated frameworks to comprehensively improve consensus performance.

To address these issues, this paper proposes a Z-number-based LSGDM method for resolving conflict behaviors in social network environments. The main contributions are summarized as follows: (1) A discrete Z-number distance measure and a trust-based K-means clustering method are developed to effectively group socially correlated DMs. (2) Discrete Z-number consistency and inconsistency indices are defined to characterize group evaluations, alongside a multi-indicator weighting method to overcome single-indicator limitations. (3) An integrated feedback mechanism combining opinion adjustment and conflict management is constructed to jointly optimize evaluations and trust networks, enhancing consensus efficiency. Additionally, a modified Z-number PROMETHEE II method is designed to simplify calculations and improve applicability.

The remainder of this paper is organized as follows: Section 2 introduces the preliminaries; Section 3 defines a novel Z-number distance measure; Section 4 constructs Z-number-based support and opposition relationships; Section 5 details the proposed LSGDM framework; Section 6 validates the method via numerical experiments; and Section 7 concludes the paper and outlines future research.

2. Preliminaries

This section introduces fundamental concepts of Z-numbers.

2.1. Z numbers

In this section, we introduce the basic concepts of Z-numbers and discrete Z-numbers.

Definition 1 [1]. Let $Z=(A,B)$ be a Z-number, where $A=\left\{\left(x,\mu_A(x)\right)\mid x\in[0,1]\right\}$ denotes the fuzzy constraint on the possible values of the random variable X, and $B=\left\{\left(b,\mu_B(b)\right)\mid b\in[0,1]\right\}$ denotes the fuzzy constraint on the reliability of the uncertain information described by A.

Definition 2 [9]. Let $Z=(A,B)$ be a Z-number; if A and B are two discrete fuzzy numbers and their membership functions are $\mu_A{:}\{x_1,x_2,...,x_n\}\to[0,1]$, $\{x_1,x_2,...,x_n\}\in R$, $\{b_1,b_2,...,b_n\}\to[0,1]$, and $\{b_1,b_2,...,b_n\}\in[0,1]$, this Z-number is called discrete Z-number.

Definition 3 [10]. Let Z_1 and Z_2 be two fuzzy numbers; then, the distance measures between Z_1 and Z_2 should satisfy the following conditions:

- $0\le D(Z_1,Z_2)\le 1$;
- $D(Z_1,Z_2)=0$ if and only if $Z_1=Z_2$;
- $D(Z_1,Z_2)=D(Z_2,Z_1)$;

- If $Z_1 \subseteq Z_2 \subseteq Z_3$, then $D(Z_1,Z_3) \geq D(Z_1,Z_2)$.

3. Novel Z-number distance measure

This section introduces the novel Z-number distance and similarity measure.

Definition 4. Let $Z_1 = (A_1,B_1), Z_2 = (A_2,B_2)$ be two discrete Z-numbers, where $\mu_{Ai}:\{x_{i1},x_{i2},x_{i3},\ldots,x_{in}\} \to [0,1], \mu_{Bi}:\{b_{i1},b_{i2},b_{i3},\ldots,b_{im}\} \to [0,1], i = 1,2$ are the membership functions of A_i, B_i; the distance measure between Z_1 and Z_2 is as follows:

$$D(Z_1,Z_2) = \sqrt{\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\frac{\left|\mu_{A1i}(x_i)-\mu_{A2i}(x_i)\right|^2}{4}+\frac{1}{m}\sum_{j=1}^{m}\frac{\left|\mu_{B1j}(b_j)-\mu_{B2j}(b_j)\right|^2}{4}\\&+\frac{1}{nm}\sum_{j=1}^{m}\sum_{i=1}^{n}\frac{\left(\max\left\{\left|\mu_{A1i}(x_i)-\mu_{A2i}(x_i)\right|,\left|\mu_{B1j}(b_j)-\mu_{B2j}(b_j)\right|\right\}_j\right)^2}{2}.\end{aligned}} \tag{1}$$

Theorem 1. Eq. (1) satisfies the four conditions of the distance formula in Definition 3.

Definition 5. Let Z_1 and Z_2 be two discrete Z-numbers; then, the similarity measure between Z_1 and Z_2 is: $S(Z_1,Z_2) = 1 - D(Z_1,Z_2)$.

4. Support and opposition relations of Z-numbers

Since the potential probability distributions $P_X(x)$ involved in Z-numbers are unknown and nonconstant, a model can be constructed to generate the probability distributions.

Definition 6. Let $Z_1 = (A_1,B_1), Z_2 = (A_2,B_2)$ be two discrete Z-numbers, where $\mu_{Ai}:\{x_{i1},x_{i2},x_{i3},\ldots,x_{in}\} \to [0,1], \mu_{Bi}:\{b_{i1},b_{i2},b_{i3},\ldots,b_{im}\} \to [0,1], i = 1,2$ are the membership functions and where $P_m(X_i)$ is the probability distribution of X. The thresholds for indifference, preference, and veto are denoted as id, p, and v, $0 \leq id < p < v < 1$. The concordance and indices between Z_1, Z_2 are as follows:

$$CI(Z_1,Z_2) = C(CZ_1,CZ_2) = \begin{cases}0, & CZ_1 + p \leq CZ_2\\ (CZ_1 - CZ_2 + p)/(p - id), & CZ_1 + id < CZ_2 < CZ_1 + p.\\ 1, & CZ_1 + id \geq CZ_2\end{cases} \tag{2}$$

$$DI(Z_1,Z_2) = D(CZ_1,CZ_2) = \begin{cases}0, & CZ_2 - CZ_1 \leq p\\ (CZ_1 - CZ_2 - p)/(v - p), & p < CZ_2 - CZ_1 < v.\\ 1, & CZ_2 - CZ_1 \geq v\end{cases} \tag{3}$$

where $CZ_i = \alpha g C_i^{AP} + \beta g C_i^{B}$, $C_i^{AP} = \left(\frac{1}{m}\right)\sum_{l=1}^{m}\frac{\sum_{k=1}^{n} x_{ik}\mu_{iA}(x_{ik})P_{lk}(x_{ik})}{\sum_{j=1}^{2}\sum_{k=1}^{n} x_{jk}\mu_{jA}(x_{jk})P_{lk}(x_{jk})}$, $C_i^{B} = \frac{\sum_{k=1}^{m} b_{ik}\mu_{iB}(b_{ik})}{\sum_{j=1}^{2}\sum_{k=1}^{m} b_{jk}\mu_{jB}(b_{jk})}$.

Definition 7. Let $Z_1=(A_1,B_1),Z_2=(A_2,B_2)$ be two discrete Z-numbers, where $CI(Z_1,Z_2),CI(Z_2,Z_1)$ are concordance indices between Z_1 and Z_2; then, the support relation between Z_1,Z_2 can be defined as:

- Strong support $(Z_1\ ?\ _{Ss}\ Z_2): CI(Z_1,Z_2)-CI(Z_2,Z_1)=1$
- Moderate support $(Z_1\ ?\ _{Ms}\ Z_2): p\le CI(Z_1,Z_2)-CI(Z_2,Z_1)<1$
- Weak support $(Z_1\ ?\ _{Ws}\ Z_2): id<CI(Z_1,Z_2)-CI(Z_2,Z_1)<p$
- Indifferent support $(Z_1\ ?\ _{Is}\ Z_2): \left|CI(Z_1,Z_2)-CI(Z_2,Z_1)\right|\le id$

Definition 8. Let $Z_1=(A_1,B_1),Z_2=(A_2,B_2)$ be two discrete Z-numbers, where $DI(Z_1,Z_2)$, $DI(Z_2,Z_1)$ are the indices between Z_1 and Z_2; then, the opposition relation between Z_1,Z_2 can be defined as:

- Strong opposition $(Z_1\ ?\ _{So}\ Z_2): DI(Z_1,Z_2)-DI(Z_2,Z_1)=1$
- Moderate opposition $(Z_1\ ?\ _{Mo}\ Z_2): v\le DI(Z_1,Z_2)-DI(Z_2,Z_1)<1$
- Weak opposition $(Z_1\ ?\ _{Wo}\ Z_2): p<DI(Z_1,Z_2)-DI(Z_2,Z_1)<v$
- Indifferent opposition $(Z_1\ ?\ _{Io}\ Z_2): \left|DI(Z_1,Z_2)-DI(Z_2,Z_1)\right|\le p$

5. Large-scale group decision-making method with Z-numbers

Let the set of DMs be denoted as $E=\{e^1,e^2,e^3,...,e^s\}, s\ge 20$ and the sets of alternatives and criteria be denoted as $Al=\{al_1,al_2,...,al_P\}$ and $At=\{at_1,at_2,...,at_Q\}$. In the process, the weights of the DMs are represented as $W^E=\{W^1,W^2,W^3,...,W^s\}$, $\sum_{T=1}^{s}W^T=1$, while the weights of the criteria are denoted as $\eta=\{\eta^1,\eta^2,\eta^3,...,\eta^Q\}$, $\eta^1,\eta^2,\eta^3,...,\eta^Q\ge 0$, and $\sum_{q=1}^{Q}\eta^q=1$. The evaluation matrix of DMs in the K-th iteration is represented as $P_T^K=\left[p_{ij}^{TK}\right]_{P\times Q}$, where p_{ij}^{TK} indicates e^T's evaluation of criterion at_j for alternative al_i.

5.1. Group clustering

This paper proposes an improved K-means clustering algorithm. This improved K-means algorithm first imports evaluation matrices, trust matrices and preset parameters, and randomly selects initial cluster centroids. Within the maximum iteration limit, it repeatedly classifies decision makers based on the integrated evaluation-trust score and updates centroids via cluster centrality metrics. It calculates the global evaluation error after each clustering assignment, terminates iterations once the error falls below the convergence threshold, and finally outputs stable clustering partitions.

5.2. Clustering computational complexity analysis

The proposed algorithm has a time complexity of $O(I\cdot N\cdot K\cdot d)$. While matching

standard K-means, it scales linearly with the number of decision makers N, significantly outperforming the $O(N^3)$ complexity of classic hierarchical clustering.

5.3. Consistency testing

In this section, we introduce a Z-number-based consistency testing method.

First, the cluster evaluations are aggregated, and the evaluation $P_{C_r} = \left[p_{pq}^{C_r} \right]_{P\times Q}$ of cluster C_r can be obtained as follows:

$$p_{pq}^{C_r} = Z_{WBM}\left(p_{pq}^1, p_{pq}^2, ..., p_{pq}^{l_r} \right) = \left(\bigoplus_{i\neq j}^{i,j\in l_r} \frac{W_r^i W_r^j}{1-W_r^i} \left(p_{pq}^{i\ h} \otimes p_{pq}^{j\ h} \right) \right)^{\frac{1}{g+h}}. \quad (4)$$

The group consensus matrix can be constructed on the basis of the evaluation similarities between clusters as follows:

$$px_{pq} = \left[\sum_{i=1}^{k-1} \sum_{j=i+1}^{k} W_{ij} S\left(p_{pq}^i, p_{pq}^j \right) \right] \Big/ \left(\sum_{i=1}^{k-1} \sum_{j=i+1}^{k} W_{ij} \right). \quad (5)$$

where $W_{ij} = \min\left(W_{C_i}, W_{C_j} \right) \in [0,1]$.

The group consensus level on criterion at_j of alternative al_i can be obtained as:

$$GPX_{pq} = px_{pq}. \quad (6)$$

The group consensus level on alternative al_i can be obtained as:

$$GPX_p = \left(\sum_{q=1}^{Q} GPX_{pq} \right) \Big/ Q. \quad (7)$$

The overall group consensus level can be obtained as:

$$GCL = \left(\sum_{p=1}^{P} GPX_p \right) \Big/ P. \quad (8)$$

5.4. Feedback mechanism

If evaluation consistency is below the threshold, a feedback mechanism provides adjustment advice for DMs. This work designs a feedback mechanism, and proposes a method to identify and resolve conflicts for higher consensus efficiency.

(1) Conflict behaviour identification

Step 1: Compute e_i's support-opposition deviation in the K-th consensus round:

$$TSO_i^K = \left(s. \sum_{j=1,j\neq i}^{s} SOI_{ij}^K \right) \Big/ \left(\sum_{i=1}^{s} \sum_{j=1,j\neq i}^{s} SOI_{ij}^K \right). \quad (9)$$

where SOI_{ij}^K represents the support-opposition index of e_i to e_j in the K-th round.

Step 2: Calculate e_i's trust in other DMs for Round K consensus:

$$tr_i^K = \left(s \cdot Ctd_K{}^i \right) \Big/ \sum_{i=1}^{s} Ctd_K{}^i. \quad (10)$$

where $Ctd_K^{\,i}$ denotes the comprehensive degree of trust of e_i in the K-th round.

Step 3: Determine conflict behaviours. In the K-th round of consensus, if the support–opposition deviation of e_i is $TSO_i^K \geq \chi_1$ and the degree of trust of e_i towards others is $tr_i^K < \delta$, e_i can be determined to exhibit conflict behavior.

(2) Conflict behaviour management

When a DM is identified as exhibiting conflict behaviours in the K-th consensus round, other DMs adjust their degrees of trust accordingly in the K+1-th round as follows:

$$t_{ji}^{K+1} = \left(1-\rho_i^K\right)t_{ji}^K, \quad \text{j=1,2,...,s , j} \neq \text{i.} \tag{11}$$

where ρ_i^K is the adjustment degree, which is graded based on the support–opposition deviation as:

$$\rho_i^K = \begin{cases} 0.2, & TSO_i^K \geq \chi_3 \text{ and } tr_i^K < \delta \\ 0.35, & \chi_2 \leq TSO_i^K < \chi_3 \text{ and } tr_i^K < \delta \\ 0.5, & \chi_1 \leq TSO_i^K < \chi_2 \text{ and } tr_i^K < \delta \\ 0, & \text{other} \end{cases}. \tag{12}$$

5.5. Alternatives ranking

Based on the defined Z-number concordance and discordance indices, an enhanced PROMETHEE II method is proposed.

The net flow can subsequently be determined as follows:

$$\varphi\left(al_i\right) = \varphi_+\left(al_i\right) - \varphi_-\left(al_i\right); \quad \text{i=1,2,...,p.} \tag{13}$$

Where $\varphi_+\left(al_i\right) = \sum_{K=1}^{s} W_K \dfrac{\sum_{j=1,j\neq i}^{p}\sum_{r=1}^{q} sri_{ij}^K\left(p_{ir}^K, p_{jr}^K\right)}{p-1}$; $\varphi_-\left(al_i\right) = W_K \dfrac{\sum_{j=1,j\neq i}^{p}\sum_{r=1}^{q} ori_{ij}^K\left(p_{ir}^K, p_{jr}^K\right)}{p-1}$.

6. Case study

A medical device company plans to select the best supplier from al_1, al_2, al_3, al_4 four suppliers. The DMs evaluate the suppliers on the basis of four criteria: at_1 price, at_2 quality, at_3 delivery time, and at_4 service level, and each criterion $\eta = \{0.2, 0.2, 0.5, 0.1\}$ is weighted accordingly. The purchasing decision-making group comprises 20 DMs who evaluate the alternatives based on the information presented [8].

Step 1: Group clustering.

The number of clusters is set to $t = 3$, and the parameters of the consensus score are $\omega = 0.5, \psi = 0.5$. The improved clustering method divides the 20 DMs into three clusters: $C_1 = \{e_4, e_6, e_8\}$, $C_2 = \{e_1, e_2, e_5, e_7, e_9, e_{10}, e_{12}, e_{13}, e_{14}, e_{16}, e_{17}, e_{18}, e_{20}\}$ and $C_3 = \{e_3, e_{11}, e_{15}, e_{19}\}$.

Step 2: Consistency testing.

From the trust relationships, the support–opposition relationships, and the of DMs, the weights of DMs and clusters are computed. Then, the support–opposition index and evaluation consistency are calculated via Eqs. (4-13).

The group consensus level of alternatives was calculated based on the group consensus matrix. Thus, the group consensus level can be calculated as 0.89945. Given that the initial threshold was set at 0.95, the current consensus level is below this threshold. So, it is necessary to engage the feedback mechanism to adjust the information.

Step 3: Feedback mechanism.

First, generate adjustment suggestions to update evaluations and recalculate support–opposition indices. Compute DMs' support–opposition deviation and trust degrees to identify conflict-prone DMs.

Step 4: Alternatives ranking.

The PROMETHEE II net-flow results indicate that al_4 has the highest net flow 0.1384, followed by al_1 (0.0472) and al_2 (0.0306), whereas al_3 has the lowest net flow (-0.2685). Therefore, the final ranking is $al_4 \succ al_1 \succ al_2 \succ al_3$.

7. Conclusion

This paper develops a Z-number-based LSGDM method addressing group conflicts. A Z-number distance optimizes K-means clustering; concordance indices combining support-opposition relations enable multi-index weighting. A feedback mechanism handles evaluation adjustment and conflicts, and PROMETHEE II is upgraded using these indices. A supplier selection case with sensitivity and comparison analyses proves the method's validity and advantages. Future work includes: (1) quantifying criterion interactions via fuzzy measures; (2) extending to multi-stage decisions using dynamic networks; (3) integrating temporal trust evolution in LSGDM.

References

[1] Zadeh LA. A Note on Z-numbers. *Information Sciences,* 2011,181:2923–32.
[2] Cao M, Gai T, Wu J, Chiclana F, Zhang Z, Dong Y, et al. Social network group decision making: Characterization, taxonomy, challenges and future directions from an AI and LLMs perspective. Information Fusion, 2025,121:103107.
[3] Wang Y H, Ding R X, Ju Y, Xu Y, Chiclana F. Three-way decision-based consensus feedback mechanism considering conflict level and social network for large-scale group decision-making. Information Fusion, 2026,126:103634.
[4] Li J, Xiang Y, Peng H. An extended multi-criteria group decision-making method based on preference ranking under Z-number environments. Engineering Applications of Artificial Intelligence, 2025,149:110573.
[5] Liao H, Wu Z, Tang M, Wan Z. An interactive consensus reaching model with updated weights of clusters in large-scale group decision making. *Engineering Applications of Artificial Intelligence*, 2022,107:104532.
[6] Rodríguez RM, Labella Á, Sesma-Sara M, Bustince H, Martínez L. A cohesion-driven consensus reaching process for large scale group decision making under a hesitant fuzzy linguistic term sets environment. *Computers & Industrial Engineering*, 2021,155:107158.
[7] Lang N N, Wang L, Zha Q B. Network game in group decision making: managing consensus with incentive and interaction interventions. *European Journal of Operational Research*, 2026, 329(3), 950-965.
[8] Peng H, Wang X, Zhang H, Wang J. Group decision-making based on the aggregation of Z-numbers with Archimedean t-norms and t-conorms. *Information Sciences,* 2021, 569:264–86.
[9] Aliev RA, Alizadeh AV, Huseynov OH. The arithmetic of discrete Z-numbers. Information Sciences, 2015, 290:134–55.
[10] Voxman W. Some remarks on distances between fuzzy numbers. Fuzzy Sets and Systems, 1998, 100:353–65.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)

© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261140

Carotid Plaque Prediction Based on Multi-Channel PPG Signals

Shuo Yuan [a], Guozhong Liu [a*], Hengli Liu [a] and Hao Meng [a]
[a] *Beijing Information Science and Technology University, Beijing 100192, China*

Abstract. This study proposes a non-invasive carotid plaque prediction method using multi-channel photoplethysmography (PPG) signals. A six-channel synchronous acquisition system records signals from bilateral carotid, superficial temporal, and radial arteries. Features are extracted from five dimensions—time-domain, morphological, derivative, inter-channel correlation, and derived physiological parameters—together with clinical indicators. Random Forest with out-of-bag permutation importance selects the top 20 features from the original 203, followed by a linear Support Vector Machine (SVM) classifier evaluated via five-fold cross-validation. The model attains 88.06% accuracy, 0.95 recall, 0.89 F1-score and an AUC of 0.918, confirming that multi-channel PPG integrated with machine learning performs well in non-invasive carotid plaque risk screening.

Keywords. Photoplethysmography, Carotid Plaque, Feature Selection, Random Forest, Support Vector Machine, Machine Learning

1. Introduction

Carotid plaque, a hallmark of atherosclerosis, constitutes a major risk factor for ischemic stroke and coronary artery disease. In China, cardiovascular disease prevalence is predicted to climb from 886.94 to 1151.61 per 100,000 people during 2020–2030[1], highlighting the clinical significance of early plaque detection. Current clinical methods—carotid ultrasound, computed tomography angiography (CTA), and magnetic resonance angiography (MRA)—face notable limitations. Ultrasound is operator-dependent and unsuitable for continuous monitoring, while CTA and MRA involve radiation exposure, contrast agent risks, long examination times, or high costs, limiting their use in large-scale or home-based screening. Previous work has established PPG feature extraction frameworks [2-4], demonstrated the value of multi-site measurements [5], and applied machine learning to vascular aging prediction [6]. However, most existing studies rely on single-site measurements or limited feature sets, lacking multi-channel fusion exploration for carotid plaque prediction. To bridge this gap, this study integrates multi-channel PPG signals with signal processing and machine learning to build a non-invasive carotid plaque prediction model for early cardiovascular risk screening.

[*] Corresponding Author: Guozhong Liu, Email: liuguozhong@bistu.edu.cn

2. Methods

2.1. Signal Acquisition System and Dataset Construction

The multi-channel PPG signal acquisition system adopts a "host–slave" architecture. Its slave unit built around a 16-bit microcontroller integrates six 515 nm green reflective PPG sensors to synchronously capture signals from bilateral carotid, temporal and radial arteries (Figure 1). Equipped with a 12-bit ADC, the system operates at a sampling frequency of 210 Hz (f_s), complying with the Nyquist sampling theorem.

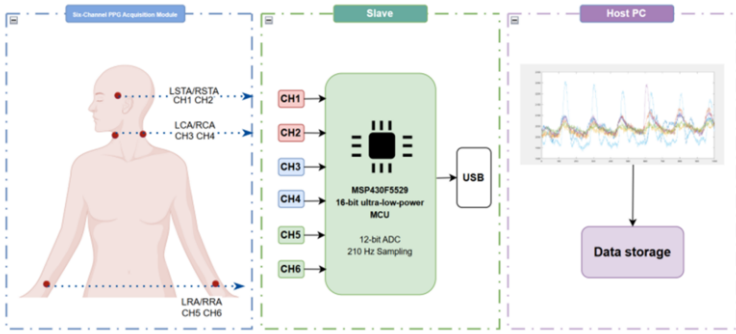


Figure 1. Schematic of six-channel PPG acquisition setup. LCA: Left Carotid Artery; RCA: Right Carotid Artery; LSTA: Left Superficial Temporal Artery; RSTA: Right Superficial Temporal Artery; LRA: Left Radial Artery; RRA: Right Radial Artery

The host unit, developed in MATLAB, provides a GUI for communication, display, and storage. The slave uses interrupt-driven ADC sampling, ensuring synchronous six-channel acquisition.

A total of 42 volunteers (14 male, 28 female, age 22–77) were recruited. All subjects underwent carotid ultrasound prior to acquisition, yielding 22 plaque cases and 20 healthy controls. Subjects rested for 5 minutes at 22–25°C before 3 minutes signal recording. Raw data were stored as text files, with 30,000 consecutive samples (≈143 s) extracted per channel. Eight clinical indicators were recorded: height, weight, diabetes history, systolic/diastolic blood pressure, smoking history, carotid-to-temporal distance, and carotid-to-radial distance.

2.2. Signal Preprocessing and Feature Extraction

Raw PPG signals contain various noise components including high-frequency myoelectric interference, power-line noise, and low-frequency baseline drift. To suppress noise while preserving waveform morphological characteristics, a Savitzky–Golay smoothing filter was applied to each channel with polynomial order 3 and window length 11 samples. This filter achieves smoothing by fitting low-order polynomials within local windows, better preserving peak and valley morphologies compared to linear filters. After filtering, the mean of each channel was subtracted to eliminate DC offset. Figure 2 shows the six-channel waveforms [7] before and after preprocessing.

Fiducial point detection followed the B→A→D→C strategy: locate systolic peak B via peak detection (minimum distance: 0.4×*fs*; minimum height: 0.5σ), then search

backward for onset A (B−0.2×*fs* to B), forward for diastolic peak D (B+0.2×*fs* to B+0.6×*fs*), and between B and D for dicrotic notch C. Invalid cycles ($t_B \leqslant t_A$ or $t_D \leqslant t_B$) were excluded. Figure 3 illustrates the four points on a typical PPG waveform.

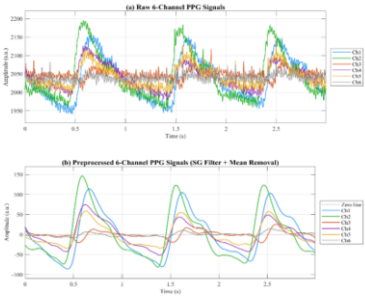


Figure 2. Waveforms before and after preprocessing.

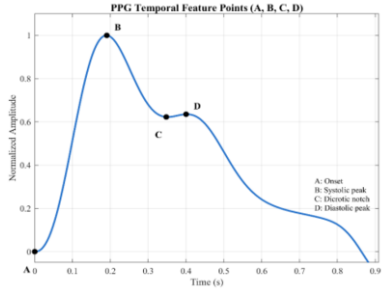


Figure 3. Fiducial point annotation.

Feature extraction was performed based on the detected fiducial points across five dimensions, as summarized in Table 1.

Table 1. Summary of extracted PPG features.

No.	Time-domain	Width features	Area features	Derivative features	Physiological parameters
1	$IPR = \frac{60}{t_D - t_A}$	$SW_p = t_B - t_{\text{rise}}(p)$	$IPA = \frac{A_2}{A_1}$	$ms = \max(\text{diff}(x) \cdot f_s)$	$RI = \frac{y_D}{y_B}$
2	$amp = y_B - y_A$	$DW_p = t_{\text{fall}}(p) - t_B$		$AGI = \frac{b - c - d - e}{a}$	$AIx = \frac{y_D - y_B}{y_B}$
3	$CT = t_B - t_A$	$W_p = t_{\text{fall}}(p) - t_{\text{rise}}(p)$		$AGI_{\text{mod}} = \frac{b - c - d}{a}$	$SI = \frac{\text{height}}{t_D - t_B}$
4	$DT = t_D - t_B$			$AGI_{\text{inf}} = \frac{b - e}{a}$	
5	$T_{\text{cycle}} = t_D - t_A$				
6	$\text{duty}_{\text{sys}} = \frac{CT}{T_{\text{cycle}}}$				

Time-domain basic parameters were calculated from the times and amplitudes of the fiducial points [8]

Width features were computed at 25%, 50%, and 75% amplitude levels. For each level p, the upstroke (A to B) and downstroke (B to D) were linearly interpolated to obtain the crossing times $t_{rise(p)}$ and $t_{fall(p)}$ at threshold $p \cdot (y_B - y_A) + y_A$. SW_p is systolic width, DW_p is diastolic width, and W_p is total width.

Area features were computed using the trapezoidal method: A_1 is the area under the curve from A to C, A_2 is the area from C to the end of the cycle, and IPA is the ratio of the two areas.

First- and second-derivative features were also extracted. The first derivative reflects the rate of change of the PPG signal. For discretely sampled signals, the first derivative was obtained by multiplying the difference between adjacent samples by the sampling frequency. The maximum value was recorded as ms. This metric corresponds to the steepest point of the waveform upstroke [9] and is positively correlated with

ventricular ejection strength. In the second-derivative waveform, five characteristic peaks (a, b, c, d, e) correspond to curvature changes at different stages of the PPG waveform. Where AGI (Aging Index) and its variants are aging gradient indices used to quantify the degree of vascular wall elasticity degradation. Additionally, t_{bc} and t_{bd} were defined as the time intervals from peak b to peak c and from peak b to peak d in the second-derivative waveform (in seconds), reflecting the elastic recoil speed of the arterial wall.

Derived physiological parameters were computed from the basic features, with clear clinical physiological significance [3].

Pearson correlation coefficients between all six channel pairs yielded 15 inter-channel features. Combined with 8 clinical features, the initial set had 209 dimensions (31×6 + 8 + 15). Variance filtering (threshold < 10^{-6}) reduced this to 203 dimensions.

2.3. Feature Selection and Prediction Model Construction

Feature contributions were ranked using the out-of-bag permutation importance of the Random Forest algorithm, and only the top 20 features were retained for subsequent analysis. Z-score standardization was performed to eliminate dimensional disparities among features (e.g., SBP ~120 mmHg, IPR ~70 bpm, width features ~0.1 s).

$$x_{\text{norm}} = \frac{x - \mu}{\sigma} \tag{1}$$

A linear-kernel support vector machine (SVM) classifier is employed for the 20-dimensional filtered features owing to its low parameter complexity and good interpretability. The penalty parameter C was tuned via grid search over {0.01, 0.1, 1, 10, 100} using five-fold cross-validation, with average accuracy as the selection criterion. The 42 samples were randomly partitioned into 5 folds, each fold is alternately used as the test set with the rest for training, and the averaged result across five runs is adopted as the final performance.

3. Results

3.1. Feature Selection Results

The top 20 features ranked by Random Forest out-of-bag (OOB) permutation importance are presented in Table 2.

Table 2. Top 20 features ranked by importance.

Rank	Feature	Channel	Physiological Significance
1	Ch4_CT	Right carotid artery	Systolic time, reflecting rapid ventricular ejection
2	Ch4_DW25	Right carotid artery	25% diastolic width, reflecting peripheral resistance
3	Ch1_ms	Left carotid artery	Maximum slope, reflecting ventricular ejection strength
4	Ch1_DT	Left carotid artery	Diastolic time
5	Ch1_DW25	Left carotid artery	25% diastolic width
6	SBP	Clinical	Systolic blood pressure

Rank	Feature	Channel	Physiological Significance
7	Ch2_DT	Left superficial temporal artery	Diastolic time
8	Weight	Clinical	Body weight (kg)
9	Ch3_SI	Right superficial temporal artery	Stiffness index
10	Ch1_SI	Left carotid artery	Stiffness index
11	Ch2_DW25	Left superficial temporal artery	25% diastolic width
12	Ch4_duty_sys	Right carotid artery	Systolic duty ratio (systolic time/cycle duration)
13	Ch1_A1	Left carotid artery	Systolic area under the curve
14	Ch2_SI	Left superficial temporal artery	Stiffness index
15	Ch1_T	Left carotid artery	Cardiac cycle duration
16	Ch1_IPR	Left carotid artery	Instantaneous pulse rate (bpm)
17	Ch4_c_a	Right carotid artery	Second derivative c/a amplitude ratio
18	DBP	Clinical	Diastolic blood pressure
19	Ch4_DW50	Right carotid artery	50% diastolic width
20	Ch5_DW50	Left radial artery	50% diastolic width

Carotid channels dominated the top 20 (12 features), followed by temporal (4), clinical (3: SBP, weight, DBP), and radial (1). DW25 and SI appeared most frequently (3 times each), with DT and DW50 appearing twice and CT once, indicating their strong discriminative power for plaque classification.

3.2. Model Performance Evaluation

Grid search determined the optimal penalty coefficient C = 10. Five-fold cross-validation was used to comprehensively evaluate the linear kernel SVM model. Table 3 and Table 4 summarize the accuracy of each fold and the averaged evaluation metrics, respectively.

Table 3. Accuracy of Each Fold in 5-Fold Cross Validation

Fold	1	2	3	4	5
Accuracy	75.00%	100.00%	77.78%	87.50%	100.00%

Table 4. Average performance metrics

Metric	Average Accuracy	Precision	Recall	F1-Score
Value	88.06%	0.85	0.95	0.89

The model achieved an average accuracy of 88.06%. The recall (0.95) was significantly higher than precision (0.85), indicating high detection sensitivity for plaque-positive samples, i.e., a low false-negative rate. In screening scenarios, the cost of missing high-risk individuals is far greater than the subsequent examination burden of false positives, making this characteristic practically significant for early screening. Per-fold accuracies ranged from 75% to 100%, reflecting the impact of data partitioning on model performance under small-sample conditions, but the five-fold average remained at a high level, indicating promising preliminary performance.

Figure 4 presents the Receiver Operating Characteristic (ROC) curve of the model, with an area under the curve (AUC) of 0.918, indicating that the classifier possesses good discriminative ability to effectively distinguish between the carotid plaque group and the healthy control group.

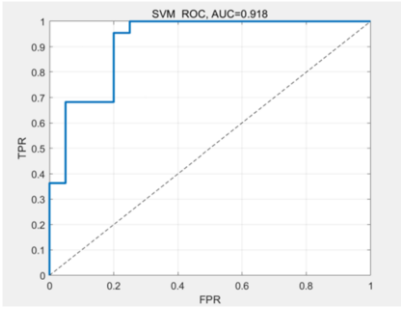


Figure 4. ROC curve of the linear SVM model (AUC = 0.918).

4. Discussion

- Physiological Interpretability of Feature Selection: Systolic time (CT), 25% diastolic width (DW25), and stiffness index (SI) were identified as the most discriminative features. CT prolongation reflects increased ventricular afterload due to reduced vascular elasticity, which explains why CT from Channel 4 ranked first. DW25 captures the elastic recoil capacity during diastolic blood return: stiffer vessels exhibit steeper downslopes, resulting in reduced DW25. SI, calculated as height divided by pulse transit time, directly reflects arterial stiffness and ranked high across two channels. The inclusion of systolic blood pressure, a well-established cardiovascular risk factor, further supports the clinical plausibility of the selected features.
- Effectiveness of Multi-Channel Fusion: The six-channel strategy covers target vessels and peripheral references. All top 5 features originated from carotid channels, confirming that target-vessel PPG signals carry the most direct plaque-related information. Features from temporal and radial channels also appeared in the top 20, providing supplementary hemodynamic information. This complementarity enables the model to capture both local pathology and systemic vascular status—unattainable with single-site acquisition.
- Comparison with Related Studies: The proposed model achieved 88.06% accuracy and an AUC of 0.918 on 42 samples. Compared with a blood-biomarker-based Random Forest model for predicting vulnerable carotid plaques on 1,701 patients (AUC = 0.847)[10], our model achieved a higher AUC despite the smaller dataset. Ferizoli et al. reported significant correlations between PPG morphological features and arterial stiffness ($p < 0.05$)[11]; the present study validates their utility in actual plaque classification.
- Limitations: Three limitations of this work are noted: (1) The dataset contains only 42 subjects, resulting in insufficient statistical power; further validation on larger patient cohorts is necessary. (2) Fiducial point extraction depends on manually fixed thresholds such as $0.4 \times fs$, which may degrade feature quality for waveforms with drastic fluctuations. (3) the current binary classification does not account for plaque burden, number, and stability.

5. Conclusion

This paper proposed a non-invasive carotid plaque prediction method using multi-channel PPG signals. A six-channel acquisition system was developed to cover the bilateral carotid, temporal, and radial arteries, enabling simultaneous recording from target and peripheral vessels. A 203-dimensional feature set was extracted from five dimensions and reduced to 20 core features via Random Forest OOB permutation importance. Carotid artery features dominated the ranking, with DW25 and SI being the most discriminative indicators. A linear SVM classifier achieved an average accuracy of 88.06%, recall of 0.95, and AUC of 0.918. The recall significantly exceeded precision, confirming the model's suitability for screening contexts where minimizing false negatives is critical. These results demonstrate the effectiveness of multi-channel PPG combined with Random Forest feature selection and SVM for non-invasive plaque prediction. The method is low-cost, simple to operate, and radiation-free, making it promising for community-based screening. Future work includes dataset expansion, deep learning integration, and algorithm deployment on mobile platforms for real-time assessment.

References

[1] Sheng BH, Zhang XY, Yang HX, Wang YG. Trends and projections of cardiovascular disease and chronic respiratory disease burden in Chinese residents from 1990 to 2019. Zhonghua Gong Gong Wei Sheng. 2024;40(4):477-483. Chinese. doi:10.11847/zgggws1143296.

[2] Tian ZW, Jia YF. Feature extraction algorithm for cardiovascular disease-related signs based on photoplethysmography. Chuanganqi Jishu yu Yingyong. 2023;11(1):20-27. Chinese.

[3] Charlton PH, Paliakaite B, Pilt K, Bachler M, Zanelli S, Kulin D, et al. Assessing hemodynamics from the photoplethysmogram to gain insights into vascular age: a review from VascAgeNet. Am J Physiol Heart Circ Physiol. 2022;322(4):H493-H522. doi:10.1152/ajpheart.00392.2021.

[4] Park J, Shin H. Vascular aging estimation based on artificial neural network using photoplethysmogram waveform decomposition: retrospective cohort study. JMIR Med Inform. 2022;10(3):e33439. doi:10.2196/33439.

[5] Allen J, O'Sullivan J, Stansby G, et al. Age-related changes in pulse risetime measured by multi-site photoplethysmography. Physiol Meas. 2020;41(7):074001. doi:10.1088/1361-6579/ab9f4a.

[6] Shin H. XGBoost regression of the most significant photoplethysmogram features for assessing vascular aging. Sci Rep. 2022;12:10765. doi:10.1038/s41598-022-15240-4.

[7] Pal R, Rudas A, Williams T, et al. Feature extraction tool using temporal landmarks in arterial blood pressure and photoplethysmography waveforms. NPJ Cardiovasc Health. 2025;2(1):57. doi:10.1038/s44325-025-00057-0.

[8] Suboh MZ, Jaafar R, Nayan NA, Harun NH, Mohamad MSF. Analysis on four derivative waveforms of photoplethysmogram (PPG) for fiducial point detection. Front Public Health. 2022;10:920946. doi:10.3389/fpubh.2022.920946.

[9] Takazawa K, Tanaka N, Fujita M, et al. Assessment of vasoactive agents and vascular aging by the second derivative of photoplethysmogram waveform. Hypertension. 1998;32(2):365-370. doi:10.1161/01.HYP.32.2.365.

[10] E Y, Yao Z, Ge M, Huo G, Huang J, Tang Y, et al. Development and validation of a machine learning model for predicting vulnerable carotid plaques using routine blood biomarkers and derived indicators: insights into sex-related risk patterns. Cardiovasc Diabetol. 2025;24:326. doi:10.1186/s12933-025-02867-6.

[11] Ferizoli R, Karimpour P, May JM, Kyriacou PA. Arterial stiffness assessment using PPG feature extraction and significance testing in an in vitro cardiovascular system. Sci Rep. 2024;14(1):2024. doi:10.1038/s41598-024-51395-y.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261141

Wavefield Decomposition-Based Reverse Time Migration for Mirror Artifact Suppression in Horizontal Borehole Acoustic Imaging

Yonglong Yang[a1], Haitao Yan[a], Zhao Le[a], Kaifeng Wu[a] and Zhongsheng Yang[a]
[a]*CCCC Second Highway Consultants Co., Ltd., Wuhan, Hubei, China*

Abstract: Acoustic remote detection of horizontal boreholes is a crucial method for fine off-borehole geological detection in tunnel engineering under complex geological conditions. Nevertheless, conventional reverse time migration (RTM) imaging suffers from borehole-symmetric mirror artifacts, which degrade imaging accuracy and hinder practical engineering application. This paper proposes a wavefield decomposition-based imaging method for horizontal borehole acoustic remote detection. The method first constructs an analytic wavefield via the Hilbert transform and separates upgoing and downgoing wavefields in the wavenumber domain. Combined with local window wavefield decomposition and optimized imaging conditions, the proposed method suppresses borehole-side mirror interference efficiently. Different from global wavefield processing schemes with heavy computational cost, the local window decomposition strategy reduces computational complexity and improves calculation efficiency, balancing high imaging accuracy and operational performance. Synthetic tests show that the proposed method effectively improves imaging accuracy, with a mirror artifact suppression rate of more than 90%, accurately identifying real subsurface anomalous structures with minimal residual artifacts. Tunnel field applications verify the capability of the method in clearly imaging dissolution and fault fracture zones. This method solves the mirror artifact problem of conventional RTM with high efficiency and practicability, providing a reliable technical reference for off-borehole geological detection in complex and harsh geological environments.

Keywords: Horizontal borehole; Acoustic remote detection; Reverse time migration; Wavefield decomposition

1 Introduction

Under harsh environmental conditions such as high altitudes and cold regions, as well as complex geological settings, conventional engineering drilling not only poses significant construction difficulties but also causes considerable ecological damage due to equipment transportation and access road construction. More importantly, the geological information revealed by a single borehole is extremely limited. Particularly in tunnel engineering projects characterized by complex geological conditions and frequent geological hazards, Horizontal directional drilling is still constrained by the

[1] Corresponding author: Yonglong Yang, E-mail: Yoonlong@126.com

limitation of the peep-hole view, easily missing karst caves, underground rivers, faults, and other geological hazard bodies. This situation poses potential risks to tunnel design, construction, and long-term operation and maintenance. To effectively identify these hidden hazardous bodies adjacent to boreholes, in-hole testing techniques have been gradually introduced and applied to detect geological information outside the borehole.

However, the detection range of commonly used in-hole testing techniques in current engineering exploration (excluding cross-hole tomography) is mostly limited to the borehole wall or within 3 m outside the borehole [1-3]. For example, the detection range of conventional lateral resistivity logging is approximately 2 m, tubular wave detection ranges from 1 to 3 m, and conventional acoustic logging can only reflect geological structures within a few centimeters outside the borehole wall [4,5]. The limitations of the above methods in detection capability have prompted researchers to explore more penetrating in-hole acoustic remote detection techniques.

Since FORTIN et al. pioneered research on remote detection equipment, significant progress has been made in in-hole acoustic remote detection technology [6]. CHABOT et al. applied dip filtering to actual data from deviated wells and obtained satisfactory imaging results [7]. WANG et al. successfully extracted effective reflected waves from acoustic logging data and achieved remote detection processing [8]. CHE Xiaohua et al. and QIAO Wenxiao et al. conducted numerical simulations of high-dip angle fracture models adjacent to boreholes using the finite element method, verifying the effectiveness of this approach [9,10]. TANG addressed the limitations of monopole compressional wave remote detection by proposing a dipole source method, which utilizes dipole shear waves to improve detection range and signal-to-noise ratio. His team subsequently further investigated horizontal shear wave (SH-wave) remote detection technology [11-13]. DENG Chengxiang et al. and CHAO Yongsheng et al. addressed the problems existing in dipole reflected shear wave imaging by conducting theoretical research on three-component sensor-based and phased array-based three-dimensional SH-wave remote detection, proposing an azimuth identification method and promoting the continuous improvement of this technology [14-16].

In recent years, research on borehole acoustic remote detection has increasingly focused on advanced imaging algorithms and wavefield processing techniques[17]. For instance, several studies have applied RTM to borehole acoustic data, demonstrating its superiority in handling complex structures compared to conventional migration methods [18,19]. Meanwhile, wavefield separation techniques, including frequency-wavenumber filtering and polarization filtering, have been widely employed to extract target wave modes and suppress unwanted noise [20-22]. However, most of these efforts have been directed toward vertical or deviated boreholes, and the specific challenges of horizontal borehole configurations remain relatively underexplored. Furthermore, while RTM offers high imaging accuracy, it is inherently susceptible to the generation of mirror image artifacts-spurious reflectors symmetrically positioned with respect to the borehole-which arise from the interference of upgoing and downgoing wavefields. Although some preliminary attempts have been made to mitigate such artifacts using modified imaging conditions or pre-processing steps, a robust and systematic solution tailored to horizontal borehole acoustic remote detection is still lacking in existing literature.

Despite the substantial progress achieved by the above studies, a long-standing problem remains unresolved in horizontal borehole applications. Due to the specificity of borehole geometry and acoustic wave propagation mechanisms, the acquired wavefield often contains contributions from both real reflectors and their symmetric

spurious reflectors with respect to the borehole, leading to the so-called mirror image artifacts in the imaging results. Such artifacts cause real geological bodies and spurious images to appear simultaneously, resulting in multi-solution in geological interpretation and severely restricting the reliable application of horizontal borehole acoustic remote detection technology in engineering practice.

To address this issue, this paper proposes a horizontal borehole acoustic remote detection imaging method based on wavefield decomposition. The main novelty of this work can be summarized as follows:

First, A novel strategy is proposed that constructs an analytic wavefield via the Hilbert transform and exploits the property that the analytic wavefield contains only positive frequency components to achieve effective separation of upgoing and downgoing wavefields in the wavenumber domain, which is more robust than conventional frequency–wavenumber filtering approaches. Second, unlike previous studies that apply global wavefield separation, the proposed method performs wavefield decomposition within a local moving window, enabling adaptive handling of non-stationary wavefield variations along the borehole trajectory. Third, A dedicated imaging condition using decomposed wavefield components is adopted, which effectively suppresses mirror interference from the opposite side of the borehole and eliminates residual artifacts, thereby significantly improving the authenticity of the final imaging profile.

By separating and reconstructing wavefield components of different propagation modes, the proposed method effectively suppresses mirror interference and improves imaging authenticity and resolution, aiming to provide a more reliable technical support for fine-scale off-borehole detection in tunnel horizontal boreholes under complex environmental conditions.

2 Method

In horizontal borehole, after acoustic wave excitation, multiple propagation modes are generated within the borehole fluid and the surrounding formation, including guided waves propagating along the borehole, body waves (compressional and shear waves) radiating into the formation, and reflected waves generated at geological interfaces. Due to the borehole diameter being much smaller than the acoustic wavelength, the borehole acoustic field exhibits distinct cylindrical wavefront characteristics, and the radiated energy is approximately symmetrically distributed on both sides of the borehole. If reverse time migration is directly applied to the acquired full-wavefield data without distinguishing wavefield components of different propagation directions, reflection energy from geological interfaces on both sides of the borehole will be generated and simultaneously presented in the imaging results. Owing to the lack of azimuthal constraints inherent to the linear geometry of horizontal boreholes, real reflectors from both sides of the borehole and their mirror-symmetric spurious reflectors will produce symmetric artifacts in the imaging profile, leading to multi-solution or even erroneous geological interpretation. To overcome this problem, this paper introduces a wavefield decomposition method. By separating wavefield components with different propagation directions, such as upgoing and downgoing waves, or left-going and right-going waves, the effective reflection information from one side of the borehole is retained while the mirror interference from the opposite side is suppressed, thereby achieving artifact-free remote detection imaging.

In two-way wave reverse time migration, the source and receiver wavefields contain wavefield components from different directions. We decompose the source and receiver wavefields into the following forms:

$$s(t,\mathbf{x}) = s_u(t,\mathbf{x}) + s_d(t,\mathbf{x}) \tag{1}$$

$$r(t,\mathbf{x}) = r_u(t,\mathbf{x}) + r_d(t,\mathbf{x}) \tag{2}$$

Let $s(t,\mathbf{x})$ and $r(t,\mathbf{x})$ denote the source wavefield and the receiver wavefield, respectively. Let $s_u(t,\mathbf{x})$ and $s_d(t,\mathbf{x})$ represent the upgoing and downgoing components of the source wavefield, and $r_u(t,\mathbf{x})$ and $r_d(t,\mathbf{x})$ represent the upgoing and downgoing components of the receiver wavefield, respectively. $\mathbf{x} = (x,z)$ And $s(t,\mathbf{x}) = s(t,x,z)$, $r(t,\mathbf{x}) = r(t,x,z)$. According to the cross-correlation imaging condition, we readily obtain:

$$\begin{aligned} I(\mathbf{x}) &= \int_0^{T_{\max}} s_d(t,\mathbf{x}) r_u(t,\mathbf{x}) dt + \int_0^{T_{\max}} s_u(t,\mathbf{x}) r_d(t,\mathbf{x}) dt \\ &+ \int_0^{T_{\max}} s_d(t,\mathbf{x}) r_d(t,\mathbf{x}) dt + \int_0^{T_{\max}} s_u(t,\mathbf{x}) r_u(t,\mathbf{x}) dt \\ &= I_{du}(\mathbf{x}) + I_{ud}(\mathbf{x}) + I_{dd}(\mathbf{x}) + I_{uu}(\mathbf{x}) \end{aligned} \tag{3}$$

Because $I_{dd}(\mathbf{x})$ and $I_{uu}(\mathbf{x})$ cause mirror image artifacts, we primarily adopt $I_{du}(\mathbf{x})$ and $I_{ud}(\mathbf{x})$. Namely,

$$I(\mathbf{x}) = \int_0^{T_{\max}} [s_d(t,\mathbf{x}) r_u(t,\mathbf{x}) + s_u(t,\mathbf{x}) r_d(t,\mathbf{x})] dt \tag{4}$$

To guarantee the internal consistency of the directional imaging condition in Eq.(4), the Hilbert transform and local-window wavefield decomposition are applied symmetrically to both the source wavefield and the extrapolated receiver wavefield. Analytic wavefield construction and up-/down-going separation are performed identically for both wavefields. Four cross-correlation terms are generated under cross-correlation imaging condition: $I_{uu}(\mathbf{x})$, $I_{dd}(\mathbf{x})$, $I_{ud}(\mathbf{x})$ and $I_{du}(\mathbf{x})$. For the horizontal-borehole acquisition geometry adopted in this work, $I_{uu}(\mathbf{x})$ and $I_{dd}(\mathbf{x})$ correspond to wave propagation paths symmetric with respect to the borehole from the opposite side, which produce borehole-symmetric mirror artifacts. In contrast, the retained terms $I_{ud} + I_{du}$ represent valid reflections originating from real geological interfaces on one side of the borehole. Therefore, only these two terms are kept for imaging to suppress mirror interference. To utilize the imaging condition in equation (4), we decompose the wavefield into upgoing and downgoing components. According to the method proposed by Hu et al. (1987)[18], the wavefield can be readily decomposed into upgoing and downgoing waves as follows:

$$S_U(\omega,k_z) = \begin{cases} S(\omega,k_z) & \omega k_z > 0 \\ 0 & \omega k_z \le 0 \end{cases} \tag{5}$$

$$S_D(\omega,k_z) = \begin{cases} 0 & \omega k_z \ge 0 \\ S(\omega,k_z) & \omega k_z < 0 \end{cases} \tag{6}$$

Where $S(\omega,k_z)$ is the wavefield of $s(t,\mathbf{x})$ in the $\omega - k_z$ domain, ω is the angular frequency, and k_z is the wavenumber in the z-direction. $S_U(\omega,k_z)$ and

$S_D(\omega, k_z)$ denote the upgoing and downgoing source wavefield components in the $\omega - k_z$ domain, respectively. The receiver wavefield components can be obtained in a similar manner. From equations (5) and (6), it can be seen that decomposing the wavefield in the $\omega - k_z$ domain requires not only storing the wavefield at all time steps for every subsurface point but also performing 2D Fourier transforms in both time and space. Consequently, the storage and computational demands are extremely large, making the implementation rather difficult. If the wavefield contains only positive or negative frequencies, then the wavefield can be decomposed into upgoing and downgoing waves based on the sign of the wavenumber. Therefore, we resort to the analytic wavefield. The real part of the analytic wavefield is the wavefield itself, while its imaginary part is the Hilbert transform of the wavefield. The analytic wavefield contains only the positive frequency components of the original wavefield. Since finite-difference wavefield extrapolation is a linear propagation operator, we directly apply the Hilbert transform to the source term to obtain the imaginary part of the analytic wavefield[19]. The relationship between the Fourier transform and the Hilbert transform is given by equation (7):

$$F_t[h_t(s(t,\mathbf{x}))] = -i\,\mathrm{sgn}(\omega) F_t[s(t,\mathbf{x})] \tag{7}$$

$$S(t,\mathbf{x}) = s(t,\mathbf{x}) + i \cdot h[s(t,\mathbf{x})] \tag{8}$$

Let $F_t[\bullet]$ denote the Fourier transform in the time domain, $h_t[\bullet]$ denote the Hilbert transform in the time domain, and $S(t,\mathbf{x})$ denote the analytic wavefield. The specific steps for wavefield decomposition are as follows: First, convert the real-valued wavefield into an analytic wavefield via the Hilbert transform. Since performing the Hilbert transform in the time domain requires storing the wavefield values at every time step, which entails a huge computational cost, this paper applies the Hilbert transform only to the source term. Second, perform the Fourier transform on the analytic wavefield. According to equations (9)-(14), obtaining the decomposed upgoing and downgoing components of the source and receiver wavefields at a given time step requires two forward Fourier transforms and four inverse Fourier transforms at that time step, i.e., six global Fourier transforms per time step, resulting in extremely high computational cost and significant time consumption. Considering that the excitation time contributes most to the imaging result, this paper performs wavefield decomposition only at the excitation time and only within a local window, i.e., only the local wavefield is decomposed. Third, after decomposing the wavefield into upgoing and downgoing components, the RTM profile can be obtained using the imaging condition in equation (4). The algorithm flowchart is shown in Fig. 1.. Shen et al. and Wang et al. have proposed similar methods for decomposing wavefields into upgoing and downgoing components.

$$S_U(t,k_z) = \begin{cases} S(t,k_z) & k_z > 0 \\ 0 & k_z \le 0 \end{cases} \tag{9}$$

$$S_D(t,k_z) = \begin{cases} 0 & k_z \ge 0 \\ S(t,k_z) & k_z < 0 \end{cases} \tag{10}$$

$$R_U(t,k_z) = \begin{cases} R(t,k_z) & k_z > 0 \\ 0 & k_z \le 0 \end{cases} \tag{11}$$

$$R_D(t,k_z)=\begin{cases}0 & k_z \geq 0\\ R(t,k_z) & k_z<0\end{cases} \tag{12}$$

$$s_{u,d}(t,\mathbf{x}) = F^{-1}[S_{U,D}(t,k_z)] \tag{13}$$

$$r_{u,d}(t,\mathbf{x}) = F^{-1}[R_{U,D}(t,k_z)] \tag{14}$$

Where $S(t,k_z)$ and $R(t,k_z)$ are the one-dimensional Fourier transforms of the source analytic wavefield $S(t,\mathbf{x})$ and the receiver analytic wavefield $R(t,\mathbf{x})$, respectively, and k_z is the wavenumber. $S_{U,D}(t,k_z)$ denote the upgoing and downgoing components of the source wavefield in the t-k domain, $R_{U,D}(t,k_z)$ similarly for the receiver wavefield. $s_{u,d}(t,\mathbf{x})$ and $r_{u,d}(t,\mathbf{x})$ represent the upgoing and downgoing components of the source and receiver wavefields in the t-x domain, respectively, and $F^{-1}[\bullet]$ denotes the one-dimensional inverse Fourier transform. The specific steps are as follows:

First, the Hilbert transform is applied to the time domain source and receiver wavefields respectively to construct the corresponding analytic wavefields. Second, a local sliding window is imposed on the analytic wavefields in the spatial domain, and Fourier transform is performed on the wavefield within each window. Third, the propagation direction is identified according to the sign of the vertical wavenumber, so that the down-going and up-going wave components are accurately separated. Finally, inverse Fourier transform is carried out for the separated up- and down-going wave components to reconstruct pure up-going and down-going wavefields in the spatial domain. The wavefield decomposition is completed to provide clean wavefield data for the subsequent directional imaging condition.

Equations. (9)-(14) present the workflow of local-window wavefield decomposition and reconstruction based on analytic wavefields, which achieves the transformation from analytic wavefields to the t-k_z, domain and further back to the t-x domain. This set of equations involves the following physical assumptions and numerical implementation constraints:

(1) FD wavefield extrapolation is a linear operator. Hilbert transform is applied to both source and receiver wavefields to construct their corresponding analytic wavefields. To avoid prohibitive storage costs caused by t-x Hilbert transform on full time-series wavefields, analytic wavefields are constructed synchronously during wavefield extrapolation. (2) The plane wave approximation and the assumption of slowly varying medium hold within each local moving window. (3) The discrete FFT is adopted as the numerical implementation for plane wave decomposition. Truncation of local windows induces spectral leakage, which is suppressed by applying a Hanning window in practical computation. (4) As an engineering simplification for t-k_z domain decomposition, forward and inverse FFT are only performed within local spatial windows at excitation time to reduce computational complexity. When severe velocity variations, dispersive guided waves, or dominant mode converted waves exist inside the window, the above approximations break down. Residual energy will appear after wavefield decomposition, and the suppression performance for mirror artifacts will deteriorate accordingly.

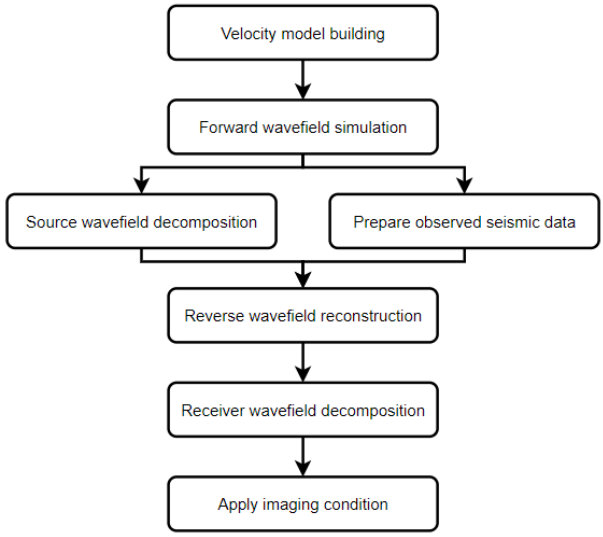


Fig. 1. Algorithm flowchart

3 Synthetic example

To validate the effectiveness of the proposed horizontal borehole acoustic imaging method based on wavefield decomposition in suppressing mirror image artifacts, a theoretical model is designed as shown in Figure 2. The model dimensions are 300 m in the x-direction and 200 m in the z-direction. The horizontal borehole is located at z=100m in the model grid and extends from the left to the right boundary. The model parameters are as follows: the P-wave velocity of the surrounding rock is 4000 m/s and the density is 2500 kg/m³; the fluid velocity inside the borehole is 1500 m/s and the density is 1000 kg/m³. Two rectangular anomalous bodies are placed on both sides of the borehole: one above the borehole (z=40m–60m, x=130m-180m) and one below the borehole (z=140m-160m, x=230m-280m). Both anomalies are characterized by a low P-wave velocity of 3000 m/s and a density of 2200 kg/m³.

The forward modeling adopts the staggered-grid finite-difference scheme for wavefield extrapolation. Absorbing boundary conditions are imposed on the outer model boundaries to mitigate artificial reflections from model truncation. Fluid-solid coupling boundary conditions are applied at the borehole wall to characterize the interaction between borehole fluid and surrounding rock and prevent extra artifacts caused by spurious boundary reflections. The source is a Ricker wavelet with a dominant frequency of 200 Hz, placed at x=50 m on the left side within the borehole. The receivers are arranged horizontally along the borehole with a trace spacing of 1 m. Conventional reverse time migration is first applied directly to the full-wavefield data obtained from forward modeling, without any wavefield decomposition. Figure 3 shows the imaging result of conventional RTM. It can be clearly observed from the figure that, in addition to the two real anomalous bodies above and below the borehole, spurious images symmetric with respect to the borehole also appear in the imaging profile: the anomalous body above the borehole produces a corresponding mirror artifact below the borehole, while the anomalous body below the borehole produces a

symmetric artifact above the borehole. These mirror artifacts are indistinguishable from the real anomalous bodies, resulting in severe ambiguity in the imaging results and making it impossible to accurately determine on which side of the borehole the anomalies are actually located. This is caused by the failure of conventional RTM to distinguish wavefield components from different propagation directions, leading to the simultaneous imaging of reflection energy from both sides of the borehole.

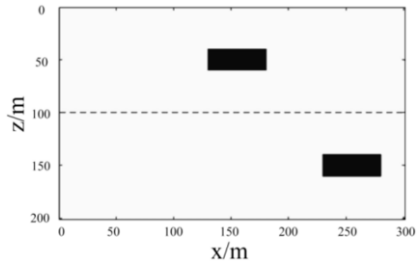


Fig. 2. Theoretical model.

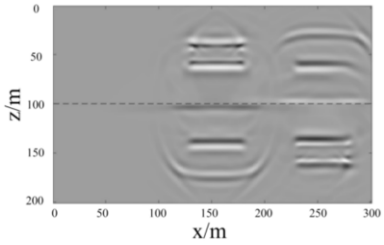


Fig. 3. Imaging profile from conventional reverse time migration.

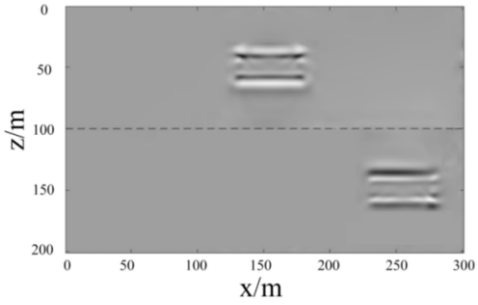


Fig. 4. Imaging profile from the improved reverse time migration.

The same dataset is processed using the proposed reverse time migration imaging method based on wavefield decomposition. First, the analytic wavefield is constructed via the Hilbert transform. Then, wavefield decomposition is performed within a local window to separate the upgoing and downgoing wavefield components. Finally, imaging is carried out using the decomposed wavefield components. Figure 4 shows the imaging profile obtained by the improved reverse time migration method proposed in this paper. It can be clearly observed that, after wavefield decomposition, the mirror artifacts on both sides of the borehole are effectively suppressed. Only the anomalous

body above the borehole and the one below the borehole are retained in the imaging profile, with each anomaly appearing only at its true position without any symmetric spurious images. Compared with Figure 3, the proposed method significantly improves the imaging authenticity and resolution, and can accurately indicate the actual spatial positions of the anomalies relative to the borehole.

To quantify mirror-artifact suppression, the residual profile is defined as $R(x,z) = I_{conv}(x,z) - I_{prop}(x,z)$, Calculations are restricted to the artifact-only region Ω to exclude energy from real geologic bodies. The artifact suppression ratio is:

$$\eta = (1 - \frac{\sum_{\Omega} \left| I_{prop} \right|^2}{\sum_{\Omega} \left| I_{conv} \right|^2}) \times 100\% \tag{15}$$

where I_{conv} and I_{prop} denote images from conventional RTM and the proposed method, respectively. $\sum_{\Omega}$ denotes summation only over the mirror-artifact zone. For the synthetic test, $\eta > 90\%$, demonstrating that more than 90 % of mirror artifacts are suppressed.

The synthetic modeling results demonstrate that, for horizontal borehole acoustic imaging, the conventional reverse time migration method cannot effectively distinguish reflection energy from the two sides of the borehole, inevitably producing symmetrically distributed mirror artifacts in the imaging results and consequently causing ambiguity in geological interpretation. In contrast, the wavefield decomposition-based imaging method proposed in this paper, by separating upgoing and downgoing wavefield components, can effectively suppress mirror interference from the opposite side of the borehole, achieve artifact-free horizontal borehole acoustic imaging, and thus accurately identify the true spatial positions of the anomalies relative to the borehole. These results validate the theoretical correctness and effectiveness of the proposed method. The extra computational overhead of the proposed method arises from analytic-wavefield construction and local-window forward/inverse FFT operations, and the local-window strategy effectively restrains the increase of computational burden. Quantitative evaluations of wall-clock time, memory consumption and model-size-dependent scalability against conventional RTM have been documented in our previous publication [19], and thus are not repeated here.

4 Field Application

To verify the applicability of the proposed method in practical engineering, an acoustic remote detection imaging test was carried out in a horizontal borehole of a tunnel. The tunnel is located in a high-altitude and cold region with complex geological conditions, where adverse geological bodies such as karst and faults are developed, posing a serious threat to tunnel construction and operational safety. The horizontal borehole in the test section has a depth of approximately 350 m and a diameter of 110 mm, penetrating marble and limestone formations. Previous geological data indicate the possible presence of geological anomalies around the borehole, such as dissolution fractures and fault fracture zones.

Field data were acquired using our self-developed in-hole acoustic remote-detection system, which has no azimuth-measurement capability. A monopole source with

dominant-frequency range of 500-1000Hz is adopted. The receiving array consists of 12 equally-spaced piezoelectric receivers with 0.5 m trace spacing. Single-point excitation with multi-channel reception is applied with 1 m excitation interval. The sampling rate is 0.02 ms and the recording length is 200 ms. The whole system is deployed along the horizontal borehole.

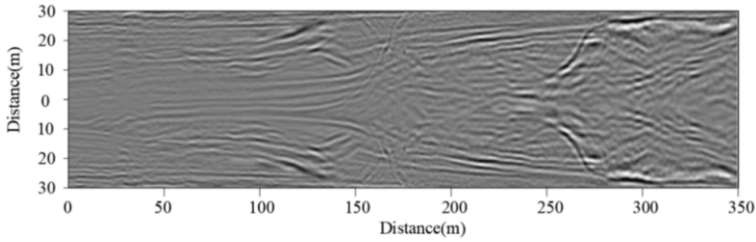


Fig. 5. Imaging profile from conventional reverse time migration.

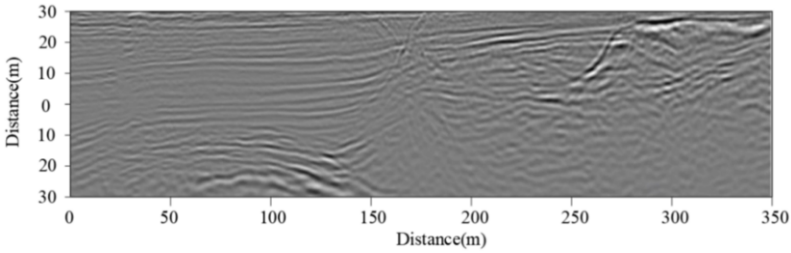


Fig. 6. Imaging profile from reverse time migration based on wavefield decomposition.

Conventional reverse time migration was first applied to the acquired data for imaging without any wavefield decomposition. Figure 5 shows the imaging profile obtained by conventional RTM. It can be observed that multiple reflective events exist in the profile. However, due to the lack of azimuthal constraints inherent to the horizontal borehole, multiple sets of mirror events symmetric with respect to the borehole appear in the imaging results. These symmetric artifacts make it difficult to determine the true spatial positions of geological interfaces and to reliably distinguish whether the anomalies are located above or below the borehole, posing considerable challenges to geological interpretation. The same dataset was processed using the proposed reverse time migration method based on wavefield decomposition, and Figure 6 shows the corresponding imaging profile. Compared with Figure 5, the mirror artifacts are effectively suppressed after wavefield decomposition, and only the effective reflection information from one side of the borehole is retained in the imaging profile, with two major anomalous bodies clearly identified.

The actual horizontal borehole acoustic detection results demonstrate that the conventional reverse time migration method is plagued by mirror artifacts, making it difficult to accurately determine the true spatial positions of geological anomalies around the borehole. The proposed wavefield decomposition-based imaging method can effectively separate upgoing and downgoing wavefield components, suppress mirror interference from the opposite side of the borehole, and achieve clear, artifact-free imaging. This provides a reliable technical means for off-borehole

geological detection in horizontal boreholes of tunnels under complex environmental conditions.

5 Discussion

This paper presents a wavefield decomposition imaging method for horizontal-borehole acoustic remote detection. Accurate separation of up- and down-going wavefields suppresses borehole geometry related mirror artifacts in conventional reverse time migration. As an extension of our prior work, the method adopts the analytic wavefield and local window Fourier decomposition framework and reconstructs a horizontal borehole oriented imaging condition without angle calculation to target mirror artifacts. Synthetic and field tests demonstrate its reliable performance for locating borehole-side anomalies under the tested conditions.

Built upon plane wave and locally slowly varying media assumptions, the decomposition accuracy deteriorates under strong borehole guided wave dispersion and severe lateral heterogeneity. Future quantitative robustness tests will assess the impacts of noise, velocity model errors and acquisition geometry perturbations to enhance method stability.

The analytic wavefield plus local window Fourier transform scheme constrains computational scale yet introduces extra Fourier operations relative to conventional RTM. Time domain filtering and deep learning driven fast decomposition will be explored to realize near real time engineering processing.

Imaging accuracy relies heavily on up-/down-going wave separation. Wave mode coupling, multiple reflections and mode conversion in real formations cannot be fully handled by scalar-based decomposition. Multi-component separation combined with P-, S- and converted wave information is expected to improve imaging fidelity for complex geology.

The algorithm is presently limited to standard horizontal borehole acquisition. Wave propagation and mirror artifact characteristics change for deviated or curved boreholes, requiring further dataset driven validation and adaptive optimization. Multi-site field experiments over interbedded soft-hard strata and strongly anisotropic formations, coupled with sensitivity tests, will evaluate its engineering robustness and applicable scope.

In conclusion, the proposed method mitigates mirror artifacts for horizontal borehole acoustic imaging and shows promising theoretical and application potential under given test conditions. Its generality should not be overstated owing to theoretical assumptions, limited test cases and dataset constraints. Future work will pursue efficiency optimization, complex wavefield processing, multi-borehole adaption and multi-scenario field validation to refine the algorithm.

6 Conclusions

A directional reverse time migration method based on analytic wavefield local-window decomposition is proposed to suppress borehole-symmetric mirror artifacts for horizontal borehole acoustic remote-sensing imaging. The Hilbert transform is applied to construct analytic wavefields for both source and extrapolated receiver wavefields. Up-/down-going wave separation is realized via local-window t-k_z decomposition. By

adopting the directional cross-correlation imaging condition, valid reflection terms $I_{ud} + I_{du}$ are retained while artifact-related terms $I_{uu} + I_{dd}$ are discarded, so as to mitigate mirror artifacts. Synthetic tests demonstrate that the proposed method can effectively suppress mirror artifacts with an artifact-suppression ratio greater than 90%. Promising imaging performance is also achieved using field data acquired from one horizontal borehole. Nevertheless, the present method has several limitations. It is built upon the plane wave approximation and the assumption of locally slowly-varying media. Dispersive borehole guided waves and strong lateral heterogeneity will degrade wave-decomposition accuracy. Additional computational overhead arises from local-window FFT operations and analytic wavefield construction compared with conventional RTM. Promising results are obtained under the tested cases in this study, while general applicability should not be overstated. Future work will focus on: (1) performing systematic numerical sensitivity and robustness tests with varying noise levels, velocity misfits and acquisition perturbations; (2) carrying out multi-well field validations under diverse geological conditions, and calibrating imaging results with borehole image logs and core data; (3) further optimizing the algorithm to reduce computational cost for practical engineering applications.

Reference

[1] QIN Z, MA X G,WU D, et al. Hybrid fast numerical simulation of multi-probe azimuthal gamma ray logging while drilling. Journal of East China University of Technology (Natural Science), 2020; 43(2): 145-151.
[2] YANG Y L, CHU J Q, WU D F, et al. Application of cross-hole seismic CT method in engineering investigation. Chinese Journal of Engineering Geophysics, 2021; 18(2): 178-185.
[3] YAN Haitao,YANG Yonglong, LIU Jiguo, et al. Research on forward modeling of seismic wavefields for acoustic remote detection in horizontal boreholes. Journal of East China University of Technology（Natural Science), 2026; 49(1):51-60.
[4] YANG D, WANG Y, GUI Z, et al. 2D acoustic equation prestack reverse-time migration based on an optimized combined compact difference scheme. Journal of Geophysics and Engineering, 2024; 21(5): 1445-1461.
[5] LI Q, ZHANG X F, WU X M. Application of geophysical logging in in situ leaching uranium mining. Journal of East China University of Technology (Natural Science), 2016; 39(1): 53-58.
[6] NIU Decheng, SU Yuanda. Adjacent borehole imaging method based on acoustic remote detection in shallow unconsolidated formations. Petroleum Drilling Techniques, 2022; 50(6): 21-27.
[7] FORTIN J P, REHBINDER N, STARON P. Reflection imaging around a well with the EVA full-waveform tool. The Log Analyst, 1991; 32:271-278.
[8] CHABORT L, HENLEY D C, BROWN R J, et al. Single-well imaging using the full waveform of an acoustic sonic [C]// SEG International Exposition and Annual Meeting.SEG: SEG-2001-0420 (2001).
[9] WANG N X, SU H, LIU W M, et al. Analysis of reflection image in sonic full waveform logging. Well Logging Technology, 1998; 22(4): 278-283.
[10] CHE X H, QIAO W X, YAN X Z. The effectivity of applying linear phased array transmitter in borehole acoustic reflection logging. Well Logging Technology, 2004; 28(2): 108-111.
[11] QIAO W X, CHE X H, LI G et al. The physical modeling of acoustic reflection image logging. Geophysical Prospecting For Petroleum. 2004; 43(3): 294-297.
[12] TANG X M, WEI Z T. Significant progress of acoustic logging technology: remote acoustic reflection imaging of a dipole acoustic system. Applied Acoustics (in Chinese), 2012; 31(1): 10-17.
[13] TANG X M. Imaging near-borehole structure using directional acoustic wave measurement. Geophysics, 2004; 69(6): 1378-1386 (2004).
[14] TANG X M, PATTERSON D J. Single-well S-wave imaging using multi component dipole acoustic-log data. Geophysics, 2009; 74(6): 211-223.
[15] DENG C X. Research on three-dimensional numerical modeling and imaging method of single-well acoustic reflection imaging logging[D]. Wuhan: China University of Geosciences (Wuhan) (2020).
[16] CHAO Y S, LI C, HE H B, et al. Research and application of azimuth dipole acoustic remote detecting

technology. Journal of Applied Acoustics, 2023; 42(2): 416 - 426.
[17] WEI Z T, TANG X M, SU Y D, et al. A new acoustic remote sensing method utilizing borehole low-frequency dipole shear wave. Chinese journal of Geophysics, 2013; 56 (10): 3572-3580.
[18] Hu L Z, MCMECHAN G A. Wave-field Transformations of Vertical Seismic Profiles. Geophysics, 1987; 52(3): 307-21.
[19] Yan Zhe, YANG Yonglong, and LIU Shaoyong. True Amplitude Angle Gathers from Reverse Time Migration by Wavefield Decomposition at Excitation Amplitude Time. Energies, 2020, 13(23): 6204.
[20] KONG Wentao, DONG Mingrong, LU Lin. Denoising Method for Reverse-Time Migration Based on Traveling Wave Separation and Angle Weighting. Journal of Changchun University of Science and Technology (Natural Science Edition), 2026; 49(1):33-42.
[21] Zhang L L, Zhao Y, Liu L, et al. A fast anisotropic decoupled operator of elastic wave propagation in the space domain for P-/S-wave-mode decomposition and angle calculation, with its application to elastic reverse time migration in a TI medium. Geophysics, 2023; 88(3): T151-T164.
[22] Zhou S H, Du Q Z, Lu W H, et al. Elastic wavefield separation in transversely isotropic media based on elastic parameter decomposition. Geophysical Prospecting for Petroleum (in Chinese), 2023; 62(4):614-622

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)

© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261142

Leakage-Aware Road Damage Detection based on GAM-YOLOv8

Jun WANG[a], Zuojun Bao[a], Wenjin Zeng[a], Chunting Qin[a], and Jing ZHOU[a,1]
[a] *School of Artificial Intelligence, Jianghan University, Wuhan, China*

Abstract. Automatic detection of road cracks, potholes, and repair traces is essential for intelligent transportation infrastructure inspection. However, road damage detectors are often affected by viewpoint variation, scale diversity, class imbalance, and highly similar images collected from continuous road sequences. To address these issues, this paper proposes a leakage-aware evaluation protocol and develops a lightweight road damage detection method that combines class rebalancing with global attention enhancement. Specifically, for the China subset of RDD2022, we first adopt a sequence-aware splitting method to reduce temporal leakage caused by randomly mixed adjacent frames, and then employ a two-stage visual duplicate screening strategy based on dHash perceptual hashing, pixel correlation, and mean squared error to identify high-confidence duplicate samples. Furthermore, we insert a Global Attention Mechanism (GAM) after the SPPF layer of YOLOv8n to enhance high-level channel-spatial feature modeling. Experiments show that the proposed GAM-YOLOv8n model achieves 88.05% mAP@0.5, 57.62% mAP@0.5:0.95, 83.79% precision, and 82.65% recall on the clean validation set, improving mAP@0.5 by 1.23 percentage points over the YOLOv8n baseline. More importantly, leakage cleaning shows that visual duplicates may overestimate performance by 0.4-1.0 percentage points, demonstrating the importance of strict leakage-aware evaluation for road damage detection.

Keywords. Road damage detection; YOLOv8; global attention mechanism; data leakage; class imbalance; RDD2022

1. Introduction

Automatic road damage detection is vital for ensuring pavement safety and infrastructure management [1-2]. While deep-learning-based YOLO detectors [3] are widely adopted due to their high inference speed, three main challenges persist in continuous traffic-scene imaging. First, road cracks are often thin and low-contrast against pavement textures, leading to missed detections. Second, severe class imbalance biases the model toward majority categories. Most importantly, continuous frame capturing introduces visual near-duplicates across loosely split training and validation sets, which systematically inflates evaluation metrics.

The core argument of this paper is that reliable, leakage-free evaluation is as critical as architectural optimization for credible model assessment. Based on this principle, this study utilizes YOLOv8n as the baseline and presents a coordinated framework combining rigorous data governance, class rebalancing, and attention-based feature enhancement. The main contributions are summarized below:

[1] Corresponding Author: Jing Zhou, zhj131@jhun.edu.cn. This work is supported by the National Undergraduate Innovation and Entrepreneurship Training Program (No. 202511072008).

- Leakage-Aware Protocol: Developed a sequence-aware split combined with dHash, pixel correlation, and MSE verification to filter 65 duplicate pairs, removing 35 leaked validation images.
- Class Rebalancing: Applied a training-only D40-aware oversampling strategy, raising the minority pothole instance proportion from 3.4% to 10.7% without altering validation distribution.
- GAM-YOLOv8n Architecture: Integrated a Global Attention Mechanism after the SPPF layer to boost high-level channel-spatial interactions while preserving a lightweight profile.

2. Related Work

Driven by deep learning, automated road damage detection has evolved from early detectors [3-5] toward advanced architectures tailored for irregular cracks and fine-scale targets. Notably, recent representative work [6] highlights that integrating targeted feature-enhancement modules into lightweight YOLO architectures significantly improves detection accuracy for subtle pavement distress. Building on these insights, anchor-free YOLOv8 [7] provides a highly effective baseline. However, despite architectural progress, data leakage remains a critical yet under-addressed threat to credible evaluation; intra-dataset visual leakage from near-duplicate images—particularly prevalent in adjacent frames of continuous vehicle-captured video—severely inflates evaluation metrics and impairs real-world generalization [8]. Although Arya et al. [2] acknowledged geographic overlaps in RDD2022, frame-level deduplication was omitted. To bridge this gap, we select lightweight YOLOv8n as our baseline and establish a reliable, leakage-free evaluation protocol on the RDD2022 China subset combining sequence-aware splitting, perceptual hashing, and pixel-level verification. Furthermore, we incorporate GAM [9] to retain global spatial-channel dependencies, preserving high-level feature details and isolating faint cracks from backgrounds.

3. Method

3.1 Overall Framework

The proposed framework comprises four sequential components: data preparation, class rebalancing, model enhancement, and leakage-aware evaluation. Specifically, the RDD2022 China subset undergoes sequence-aware splitting and a two-stage visual screening to filter near-duplicates, establishing a clean validation set Val-clean. Within the training set, minority D40_Pothole instances are oversampled four times to alleviate class scarcity. Finally, the Global Attention Mechanism (GAM) [9] is integrated after the SPPF layer of YOLOv8n to construct the enhanced GAM-YOLOv8n detector.

3.2 GAM-YOLOv8n

This study adopts YOLOv8n [10] as the baseline detector, which processes P3, P4, and P5 multi-scale features following the feature-pyramid principle [4]. To better capture elongated shapes, weak local textures, and large scale variations of road distresses, a Global Attention Mechanism (GAM) module is integrated immediately after the SPPF layer to enhance high-level semantic features. Formally, given an input intermediate feature map derived from the preceding SPPF layer, the sequential extraction process of

GAM can be formulated as follows:

$$F_2 = M_c(F_1) \otimes F_1, \quad F_3 = M_s(F_2) \otimes F_2 \tag{1}$$

where $\mathrm{M_c}$ and $\mathrm{M_s}$ denote the channel and spatial attention maps, respectively, and $\otimes$ represents the element-wise multiplication. Specifically, the channel attention submodule $\mathrm{M_c}$ utilizes a 3D permutation operation followed by a multi-layer perceptron (MLP) to model cross-channel dependencies. Subsequently, the spatial attention submodule $\mathrm{M_s}$ employs two convolutional layers to integrate spatial context and reduce information loss during feature reorganization.

GAM is selected over other attention modules due to three key advantages for road damage detection. First, its dual channel-spatial architecture captures both semantic and spatial dependencies, avoiding spatial loss in channel-only models (SE-Net, ECA-Net) and preserving global context lost in factorized designs (Coordinate Attention)—essential for tracing continuous, narrow cracks. Second, unlike heavy Transformer-based mechanisms, GAM maintains a lightweight footprint critical for real-time onboard deployment. Finally, compared to parameter-free alternatives (SimAM) [11], GAM's learnable parameters provide stronger representation capacity for faint distresses. Overall, GAM uniquely balances global context modeling, feature discrimination, and computational efficiency.

3.3 Network Loss

The proposed detector is optimized using the official Ultralytics YOLOv8 loss formulation [10:

$$L = \lambda_{box} L_{box} + \lambda_{cls} L_{cls} + \lambda_{dfl} L_{dfl} \tag{2}$$

where L_{box}, L_{cls},and L_{dfl}denote the bounding-box regression loss, classification loss, and distribution focal loss, respectively. We adopt the default YOLOv8 weights and use the coefficients λ_{box}, λ_{cls}, and λ_{dfl} to balance the contributions of the three loss terms, namely λ_{box}=7.5, λ_{cls}=0.5, and λ_{dfl}=1.5.

4. Experiment

4.1 Dataset and Class Distribution

This study uses the China subset of RDD2022, which contains two acquisition views: China_Drone and China_MotorBike. Five detection categories are considered: D00_Longitudinal, D10_Transverse, D20_Alligator, D40_Pothole, and Repair. After sequence-aware splitting, the training set contains 3397 images and the original validation set contains 876 images. After visual duplicate cleaning, Val-clean contains 841 images. Table 1 reports the instance distribution under different splits.

Table 1. Dataset splits and class-wise instance distribution

Split	Images	Instances	D00	D10	D20	D40	Repair
Train	3397	6811	3204	1860	719	229	799
Train-D40x4	3898	8587	3588	2124	917	916	1042
Val	876	1733	806	442	193	78	214
Val-clean	841	1648	748	429	185	73	213

4.2 Implementation Details

Models are trained for 100 epochs using the default Ultralytics YOLOv8 settings with the Adam optimizer. The initial learning rate is 0.001, the final learning-rate ratio is 0.01, weight decay is 0.0005, and the warmup period is 3 epochs. Training augmentation includes Mosaic, random flipping, and HSV color perturbation. Mosaic augmentation is disabled during the final 10 epochs.

The evaluation metrics include precision, recall, mAP@0.5, and mAP@0.5:0.95. mAP@0.5 measures average detection accuracy at an IoU threshold of 0.5, whereas mAP@0.5:0.95 evaluates localization quality over multiple IoU thresholds. Unless otherwise specified, this paper primarily reports results on Val-clean.

4.3 Leakage Detection Results

The two-stage visual screening identifies 65 high-confidence duplicate pairs, revealing that 35 validation images (4.0% of the original validation set) suffer from visual leakage across both China_Drone and China_MotorBike viewpoints.



Figure 1. Examples of high-confidence visual duplicate samples between the training and validation sets.

Fig. 1 illustrates representative near-duplicate pairs with annotated dHash distances and pixel correlations. Evidenced by removing these leaked samples, mAP@0.5 values across most models systematically drop by 0.4–1.0 percentage points. This performance decrease confirms that visual near-duplicates artificially inflate evaluation metrics, reinforcing the necessity of adopting Val-clean as our primary evaluation protocol to ensure credible model assessment.

4.4 Main Results on Val-Clean

Table 2 reports the detection results of different models on Val-clean. The GAM+D40 model achieves the highest mAP@0.5 of 88.05%, improving the baseline value of 86.82% by 1.23 pp. Its recall also increases from 81.92% to 82.65%. These results indicate that introducing high-level global attention on top of class rebalancing improves the ability of the detector to recognize road cracks and potholes.

Table 2. Detection results of different models on Val-clean

Method	P	R	mAP@0.5	mAP@0.5:0.95	Params	FLOPs	FPS
Baseline YOLOv8n	0.8372	0.8192	0.8682	0.5721	3.01	8.1	66
YOLOv10n	0.7907	0.7693	0.8265	0.5523	2.69	8.0	68
YOLOv11n	0.8261	0.8122	0.8683	0.5707	2.62	6.5	72
Baseline + D40x4	0.8403	0.8094	0.8685	0.5788	3.01	8.1	66
CBAM + D40x4	0.8273	0.8212	0.8717	0.5659	3.06	8.3	65
DLGA-DWF+DA-Loss + D40x4	0.8497	0.8060	0.8707	0.5671	3.84	10.0	59
SimAM + D40x4	0.8493	0.7994	0.8676	0.5610	3.01	8.2	66
GAM	0.8494	0.7721	0.8578	0.5593	3.34	8.5	63
GAM + D40x4	0.8379	0.8265	0.8805	0.5762	3.34	8.5	64

GAM + D40x4 + Inner-WIoU	0.8637	0.7942	0.8681	0.5709	3.34	8.5	64
GAM + P2 + D40x4	0.8421	0.8119	0.8705	0.5679	4.11	11.2	52

As Table 2 indicates, YOLOv8n outperforms YOLOv10n and matches YOLOv11n, leveraging dense feature representations to better capture low-contrast, elongated distresses. Crucially, ablation reveals a powerful synergy: standalone GAM suppresses imbalanced minority-class features (~5% D40 potholes), degrading recall. D40 oversampling effectively eliminates this bottleneck, enhancing localization quality and fully unlocking GAM's ability to capture global channel-spatial contexts for weak-textured cracks. Together, GAM+D40 achieves the top 88.05% mAP@0.5. In contrast, despite slightly higher precision, Inner-WIoU [12] and DLGA-DWF suffer severe recall drops, leading to critical missed detections of faint damage. Due to computational resource constraints, comparisons with YOLOv9 and RT-DETR are left for future work.

Figure 2 presents the training curves and PR curves of the GAM+D40 model. The left part shows training losses, precision, recall, and mAP changes, while the right part shows class-wise PR curves, which help assess convergence behavior and inter-class detection differences.

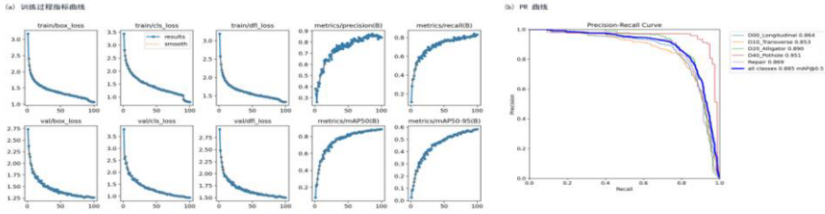


Figure 2. Training curves and PR curves of the GAM+D40 model

4.5 Statistical Validation

Table 3. Multiple Independent Training Runs with Statistical Analysis

Method	Run	P	R	mAP@0.5	mAP@0.5:0.95
Baseline YOLOv8n	3	0.8372± 0.0023	0.8192± 0.0003	0.8682± 0.0029	0.5721± 0.0031
GAM + D40x4	3	0.8379± 0.0097	0.8265± 0.0084	0.8805± 0.0046	0.5762± 0.0045

Statistical validation. Multiple independent training runs confirm the reproducibility of our results. It consistently outperforms the baseline YOLOv8n. Low standard deviations demonstrate stable training, providing empirical evidence of reliable improvement. Two-sample t-tests were performed to evaluate statistical significance. The comparison between GAM+D40 and YOLOv11n shows marginally significant improvement ($t=2.45$, $p=0.092 < 0.10$), while the comparison with baseline YOLOv8n trends toward improvement. However, effect size analysis reveals large practical significance: Cohen's $d = 1.51$ vs. baseline and $d = 2.43$ vs. YOLOv11n, indicating substantial improvements beyond mere statistical noise.

Error Analysis and Limitations. An analysis of validation failures indicates that false negatives are dominated by ultra-fine cracks losing feature resolution during downsampling, edge-truncated damage, and minority-class potholes. False positives primarily stem from complex pavement textures, shadow boundaries, and patch edges mimicking true distresses.

These errors highlight three core limitations: (1) Resolution Tradeoff: At 640×640, the model balances 64 FPS real-time speed but achieves lower AP on small targets compared to large ones; (2) Domain Sensitivity: Cross-domain performance drops on

distinct subsets validate genuine domain shift, confirming our leakage-free split; and (3) Future work will explore temporal video modeling, domain adaptation, and generative augmentation for minority distress classes.

4.6 Original Validation Set Versus Val-Clean

Table 4 compares mAP@0.5 values on the original validation set and Val-clean. After visually duplicated validation samples are removed, all models show performance decreases to varying degrees. GAM+D40 obtains 88.66% mAP@0.5 on the original validation set and 88.05% on Val-clean, corresponding to a decrease of 0.61 pp. The baseline decreases from 87.64% to 86.82%, with a reduction of 0.82 pp. These results show that model performance can be overestimated if data leakage is ignored.

Table 4. Comparison of mAP@0.5 between the original validation set and Val-clean

Method	Val mAP@0.5	Val-clean mAP@0.5	Change
Baseline YOLOv8n	0.8764	0.8682	-0.0082
Baseline	0.8752	0.8685	-0.0067
CBAM	0.8766	0.8717	-0.0049
DLGA-DWF + DA-Loss	0.8750	0.8707	-0.0042
SimAM	0.8712	0.8676	-0.0036
GAM + YOLOv8n (ours)	0.8866	0.8805	-0.0061

4.7 Ablation Study

To demonstrate the influence of evaluation protocols on experimental conclusions, this section reports early ablation results. Before data cleaning, preliminary development on the uncleaned validation set showed that integrating a P2 feature-pyramid head, CBAM, or Wise-IoU yielded minor baseline fluctuations, with Wise-IoU reaching an uncleaned mAP@0.5 of 88.52%. However, these early-stage advantages failed to maintain stability in subsequent leakage-aware sequence-split evaluations, underscoring that optimizing losses or architectures alone without data governance is insufficient for reliable road damage detection. Therefore, leakage-free evaluation and comprehensive protocol analysis are both necessary to avoid misleading architectural conclusions.

4.8 Visualization Results and Discussion

Fig. 3 illustrates representative detection results of the GAM+D40 model. While the model successfully recognizes multiple damage types in dense traffic scenes, visual analysis uncovers critical failure patterns: thin, low-contrast cracks are prone to missed detections under noisy pavement textures, whereas lane markings and repair boundaries occasionally trigger false positives. These insights clarify why other explored modules underperformed: the P2 head introduces background texture interference, CBAM/SimAM suppress low-response crack features, Inner-WIoU causes box rejection, and DLGA-DWF [13-14] increases optimization complexity. Future work will focus on segmentation-assisted supervision and hard-negative mining.

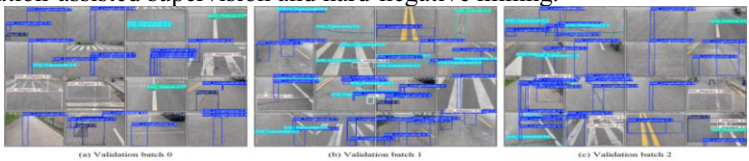


Figure 3. Road damage detection visualization results of the GAM+D40 model

5. Conclusion

This paper presents a leakage-aware GAM-YOLOv8 framework to address visual duplicate leakage, class imbalance, and low-contrast damage detection in road inspection. By incorporating sequence-aware splitting and multi-stage duplicate filtering, we construct a leak-free validation set (Val-clean). Combined with D40-aware oversampling, the proposed GAM-YOLOv8n achieves 88.05% mAP@0.5, outperforming the baseline by 1.23 pp. Crucially, our findings demonstrate that uncleaned visual duplicates cause a 0.4–1.0 pp performance overestimation, underscoring the necessity of strict leakage-aware evaluation protocols. Despite its effectiveness, this study is limited to the RDD2022 China subset for training data. Cross-dataset evaluation on India, Czech, and United States subsets reveals severe domain shift: with United_States showing notably better generalization than India and Czech. This highlights the necessity of domain adaptation techniques for practical multi-regional deployment. Additionally, this study focuses purely on object detection. Future research will extend the framework by:(1) applying unsupervised domain adaptation methods to improve cross-domain robustness, (2) exploring multi-source joint training with leakage-aware protocols to leverage diverse geographic data, and (3) investigating collaborative detection-segmentation modeling with edge-device deployment through model pruning and quantization.

References

[1] D. Arya, H. Maeda, S. K. Ghosh, D. Toshniwal, and Y. Sekimoto, "RDD2022: A multi-national image dataset for automatic road damage detection," arXiv:2209.08538, 2022.

[2] D. Arya, H. Maeda, S. K. Ghosh, D. Toshniwal, H. Omata, T. Kashiyama, and Y. Sekimoto, "Crowdsensing-based Road Damage Detection Challenge (CRDDC-2022)," arXiv:2211.11362, 2022.

[3] A. Wang, H. Chen, L. Liu, K. Chen, Z. Lin, J. Han, and G. Ding, "YOLOv10: Real-Time End-to-End Object Detection," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 37, 2024.

[4] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature Pyramid Networks for Object Detection," in Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2017, pp. 2117-2125.

[5] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and Efficient Object Detection," in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition, 2020, pp. 10778-10787.

[6] R. Zeng and C. Zhong, "YOLOv8-PD: An Improved Algorithm for Fine-Scale Pavement Distress Detection," IEEE Access, vol. 12, pp. 45120–45132, 2024.

[7] Z. Tian, Q. Zhang, and Z. Lin, "Intelligent road crack detection and analysis based on improved YOLOv8," IEEE Access, vol. 12, pp. 132450–132462, 2024.

[8] P. Ramos, R. Ramos, and N. Garcia, "Data Leakage in Visual Datasets," in Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW), 2025, pp. 1–10.

[9] Y. Liu, Z. Shao, and N. Hoffmann, "Global Attention Mechanism: Retain Information to Enhance Channel-Spatial Interactions," arXiv:2112.05561, 2021.

[10] M. Hussain, "YOLO-v1 to YOLO-v8, the Rise of YOLO and Its Complementary Nature toward Digital Manufacturing and Industrial Defect Detection," Machines, Art. 677, 2023, 11(7), doi: 10.3390/machines11070677. Available: https://doi.org/10.3390/machines11070677.

[11] L. Yang, R.-Y. Zhang, L. Li, and X. Xie, "SimAM: A Simple, Parameter-Free Attention Module for Convolutional Neural Networks," in Proc. Int. Conf. Machine Learning, 2021.

[12] Z. Tong, Y. Chen, Z. Xu, and R. Yu, "Wise-IoU: Bounding Box Regression Loss with Dynamic Focusing Mechanism," arXiv:2301.10051, 2023.

[13] Y. Qi, Y. He, X. Qi, Y. Zhang, and G. Yang, "Dynamic Snake Convolution based on Topological Geometric Constraints for Tubular Structure Segmentation," arXiv:2307.08388, 2023.

[14] J. Wang, K. Chen, R. Xu, Z. Liu, C. C. Loy, and D. Lin, "CARAFE: Content-Aware ReAssembly of FEatures," in Proc. IEEE/CVF Int. Conf. Computer Vision, 2019, pp. 3007-3016.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261143

Physics-Guided Deep Learning for Rapid 1D TEM Inversion with Uncertainty Quantification

Zhao Le, Kaifeng Wu[1], Haitao Yan, Yonglong Yang, Hui Ye, Wentao Zhang, Shaobiao Liang and Depu Hu

CCCC Second Highway Consultants Co., Ltd., Wuhan, Hubei, China

Abstract. Transient electromagnetic (TEM) methods are widely used in groundwater, mineral exploration, and engineering geophysics. However, conventional 1D TEM inversion is computationally expensive and highly non-unique. In this study, we propose a physics-guided deep learning framework that integrates forward modeling constraints, data-driven initialization, and regularized inversion to achieve both high efficiency and physical consistency. A large synthetic dataset is generated using a rigorous 1D layered-earth TEM forward operator based on frequency-domain impedance recursion and Gaver-Stehfest numerical Laplace inversion. A feedforward neural network is trained to predict initial resistivity and thickness models from logarithmic transient responses. The network output is further refined through a Tikhonov-regularized physics-based inversion. Experimental results demonstrate that the proposed hybrid approach significantly improves inversion stability and interpretability while reducing computational cost by over an order of magnitude compared with conventional methods. The framework also enables preliminary uncertainty quantification via ensemble-based prediction.

Keywords. Transient electromagnetic method; physics-guided; machine learning; hybrid inversion; uncertainty quantification.

1. Introduction

The transient electromagnetic (TEM) method is a cornerstone geophysical technique for imaging subsurface electrical resistivity structures and has become indispensable in groundwater exploration, mineral resource delineation, environmental monitoring, and near-surface engineering investigations [1-3]. By transmitting a time-varying current into the ground and recording the secondary magnetic field decay after switch-off, TEM systems achieve high sensitivity to conductive targets and offer considerable penetration depth compared with other electromagnetic methods. In particular, one-dimensional (1D) TEM inversion remains the most widely adopted interpretation routine in practice due to its computational tractability and robustness under horizontally layered geological assumptions [4].

[1] Corresponding author: Kaifeng Wu, E-mail: kaifeng_wu_cccc@163.com

Despite its popularity, quantitative interpretation of TEM data faces long-standing challenges. The underlying inverse problem is inherently nonlinear, ill-posed, and strongly affected by parameter trade-offs between layer resistivity and thickness [5]. Small perturbations in observed transients—whether caused by measurement noise, instrumental drift, or near-surface heterogeneity—can lead to markedly different inverted models. Conventional deterministic inversion algorithms, such as Occam's inversion or smoothly constrained least squares, rely on iterative local optimization schemes (e.g., Gauss-Newton or conjugate gradient methods) and require repeated evaluations of the forward operator [6, 7]. These iterations not only incur substantial computational costs but also render the inversion highly dependent on the choice of initial model. In practice, poor starting models often result in convergence to spurious local minima or physically implausible solutions, necessitating extensive manual intervention and expert interpretation [8].

In recent years, machine learning (ML), particularly deep learning, has been increasingly explored as an alternative paradigm for rapid geophysical inversion. Data-driven approaches, such as convolutional neural networks (CNNs) and fully connected networks, have demonstrated remarkable success in approximating nonlinear mappings between geophysical measurements and subsurface properties [9-11]. However, when directly applied to TEM inversion, purely data-driven models exhibit several critical limitations. First, they typically treat the inversion as a black-box regression task, largely ignoring the underlying Maxwell's equations and the physical constraints imposed by electromagnetic induction theory. As a result, ML predictions often violate basic physical laws, such as causality, energy conservation, or monotonic decay behavior, especially when extrapolating beyond the training distribution or processing noisy field data [12]. Second, the non-uniqueness of TEM inversion implies that multiple resistivity models can reproduce nearly identical transient responses; without explicit physical regularization, ML models tend to converge toward statistically averaged but geologically unrealistic solutions. Third, most existing ML inversions lack interpretability and offer no straightforward mechanism for uncertainty quantification, limiting their acceptance in mission-critical geophysical workflows.

To address these shortcomings, there has been growing interest in physics-guided or physics-informed machine learning, wherein domain knowledge is explicitly embedded into data-driven architectures [13, 14]. In the context of TEM inversion, this integration can take several forms: embedding forward modeling operators within the loss function, generating physically consistent synthetic training datasets, or coupling neural network (NN) predictions with deterministic inversion routines. Such hybrid frameworks seek to reconcile the computational expediency of ML with the physical rigor of conventional inversion, charting a viable course toward real-time and reliable subsurface characterization.

Motivated by these developments, this study proposes a novel physics-guided deep learning framework for rapid and reliable 1D TEM inversion with preliminary uncertainty quantification. Rather than replacing physics-based inversion, the NN is employed as a physics-aware initializer that provides a plausible starting model, which is subsequently refined through a Tikhonov-regularized least-squares inversion. The forward operator—formulated via frequency-domain impedance recursion and Gaver-Stehfest numerical Laplace inversion—is rigorously incorporated into both the dataset generation and refinement stages, ensuring physical consistency throughout the workflow. In addition, ensemble-based prediction is introduced to provide a first-order

estimate of epistemic uncertainty, enhancing the interpretability and robustness of the inversion results.

The principal contributions of this work are threefold:

(1) A physically constrained dataset generation strategy that embeds petrophysical priors and electromagnetic principles into synthetic data construction, effectively narrowing the solution space and suppressing non-physical predictions.

(2) A hybrid ML-physics inversion workflow that decouples large-scale parameter trend estimation (handled by a feedforward NN) from fine-scale physical refinement (handled by regularized inversion), achieving both high efficiency and stability.

(3) Improved interpretability and preliminary uncertainty quantification, enabled by physics-based regularization and ensemble prediction, which together enhance the practical utility of the proposed method for engineering and exploration applications.

The remainder of this paper is organized as follows. Section 2 describes the forward modeling approach, the construction of the physically constrained dataset, the architecture and training of the NN, and the regularized refinement procedure. Section 3 presents the experimental setup and quantitative comparison with baseline methods. Section 4 discusses the underlying mechanisms responsible for the improved performance, the implications for practical TEM surveys, and directions for future research.

2. Methodology

2.1 Forward Modeling

Accurate and stable forward modeling constitutes the physical backbone of the proposed hybrid inversion framework. In this study, the 1D TEM response is computed in the frequency domain using a recursive impedance algorithm, which is subsequently transformed into the time domain via numerical Laplace inversion. This strategy circumvents the numerical stiffness often encountered in direct time-domain solutions and maintains high precision across both early- and late-time channels.

For an N-layer earth, the wave impedance of each layer is expressed as:

$$Z_i = \sqrt{\mu_i \rho_i s} \tag{1}$$

where ρ_i and h_i denote the resistivity and thickness of the i-th layer, respectively, μ_i denotes the magnetic permeability of the i-th layer, which is assumed equal to the free-space value μ_0 throughout this study, and s is the Laplace variable. Although displacement currents are generally negligible at the low frequencies typical of TEM surveys, retaining the full Maxwellian form ensures numerical generality, particularly for early-time high-frequency components.

The total surface impedance is obtained recursively from the bottom layer upward:

$$Z_i = Z_{i+1} \frac{Z_i \tanh(\gamma_i h_i) + Z_{i+1}}{Z_i + Z_{i+1} \tanh(\gamma_i h_i)} \tag{2}$$

where $\gamma_i = \sqrt{s / \rho_i}$ is the propagation constant. This recursion implicitly satisfies the continuity of tangential electromagnetic fields at layer interfaces and naturally incorporates the influence of both conductive and inductive effects.

The time-domain transient electromagnetic response is obtained from the Laplace-domain solution using the Gaver-Stehfest numerical inversion algorithm:

$$f(t) = \frac{\ln 2}{t} \sum_{k=1}^{N} gkF(\frac{k \ln 2}{t}) \tag{3}$$

where N is the order of the algorithm (typically set to 12 or 16 for convergence), $F(s)$ is the Laplace transform of the response, and gk are the Gaver-Stehfest weighting coefficients, which are pre-computed constants determined by the algorithm's combinatorial formula. This forward operator not only generates the synthetic training data but also serves as the kernel of the physics-based refinement stage, ensuring that all inverted models remain consistent with Maxwell's equations throughout the workflow.

2.2 Dataset Generation with Physical Constraints

A fundamental weakness of purely data-driven inversion lies in its susceptibility to non-uniqueness and geologically implausible outputs. To address this, we construct a synthetic dataset under strict petrophysical and electromagnetic constraints, effectively narrowing the solution space.

Model parameters are sampled within geophysically realistic bounds, as summarized in Table 1. To further suppress unphysical extremes, two additional constraints are imposed during random sampling:

(1) Controlled resistivity contrast: Adjacent-layer resistivity ratios are capped to prevent abrupt discontinuities that rarely occur in natural sedimentary sequences. Extreme contrasts exceeding two orders of magnitude are systematically excluded.

(2) Depth-of-investigation consistency: The cumulative thickness of the model is constrained not to exceed the effective investigation depth, which is empirically inferred from the late-time decay rate of the transient response. Models in which the late-time signal falls below the expected noise floor are discarded.

Table 1. Parameter ranges for synthetic three-layer models.

Parameter	Range
ρ_1	10-200 Ω·m
ρ_2	1-500 Ω·m
ρ_3	10-1000 Ω·m
h_1	10-80 m
h_2	20-150 m

To verify that the 6000 synthetic models adhere to geological statistics (rather than being random noise), we analyze the distribution of late-time apparent resistivity values, as shown in Figure 1.

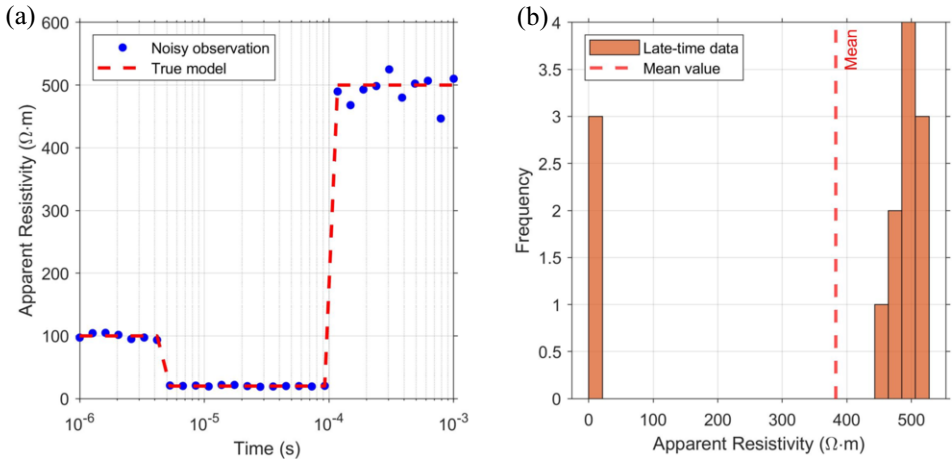


Figure 1. Distribution of late-time apparent resistivity from the synthetic TEM dataset. (a) Apparent resistivity versus time; blue dots represent noisy synthetic observations, and the red dashed line indicates the true three-layer models. (b) Histogram of late-time resistivity values, exhibiting a unimodal distribution centered near the true lower-layer resistivity (~500 Ω·m). The prominent peak near 500 Ω·m arises because late-time TEM signals are predominantly sensitive to the lowermost high-resistivity stratum set in our three-layer synthetic models. The absence of significant outliers confirms that the dataset adheres to electromagnetic principles and realistically mimics field measurement conditions, thereby supporting reliable neural network training.

The bottom layer resistivity of all synthetic three-layer models is constrained within 10-1000 Ω·m, with most samples assigned a high resistivity value close to 500 Ω·m. Since late-time transient electromagnetic responses mainly reflect the electrical property of deep formations, the statistical histogram of late-time apparent resistivity presents a distinct peak around 500 Ω·m.

Each qualified model is forward-modeled for noise-free transient response. To emulate field acquisition conditions, zero-mean Gaussian noise is superimposed according to:

$$d_{obs} = d_{syn} + \varepsilon, \varepsilon \sim \mathrm{N}(0, \sigma^2) \tag{4}$$

The signal-to-noise ratio is set to 30 dB to reflect typical near-surface TEM surveys, accounting for ambient electromagnetic interference and instrument noise. Finally, a logarithmic transformation is applied to the input data:

$$\widetilde{d} = \log_{10}(d_{obs}) \tag{5}$$

This transformation compresses the wide dynamic range of the TEM decay curve, preserves relative amplitude variations across decades of time, and substantially stabilizes NN training by mitigating the dominance of early-time high-amplitude channels.

2.3 Neural Network as Initial Solver

Rather than functioning as a standalone black-box inverter, the NN is deliberately positioned as a physics-aware initializer. Its role is to rapidly approximate a geologically plausible starting model, which is subsequently refined through deterministic inversion.

The network adopts a fully connected feedforward architecture. The input vector consists of logarithmically transformed transient responses, while the output vector

comprises five target parameters $[\rho_1, \rho_2, \rho_3, h_1, h_2]$. The architecture includes two hidden layers with 128 and 64 neurons, respectively, employing hyperbolic tangent activation functions to accommodate both positive and negative feature mappings. Training is conducted using the Levenberg-Marquardt algorithm, which interpolates between Gauss-Newton and steepest descent depending on the damping parameter. The loss function is defined as:

$$\phi = \frac{1}{N}\sum (m_{pred} - m_{true})^2 \tag{6}$$

where N is the number of training samples.

Key training hyperparameters are set as follows: the initial learning rate is fixed to 0.001; the batch size for each training iteration is set to 256; the maximum training epoch is limited to 200, and early stopping is triggered when the validation loss does not decline within 20 consecutive epochs to avoid overfitting.

The dataset is partitioned into training, validation, and test subsets at an 80:15:5 ratio. Early stopping is implemented to terminate training if the validation loss fails to improve over 20 consecutive epochs, thereby mitigating overfitting. The learning curves of the training and validation losses, along with the early stopping trigger, are presented in Figure 2. Crucially, the network is not expected to produce a final inverted model; instead, it learns to infer large-scale resistivity trends and approximate layer boundaries, providing a high-quality initial guess that significantly reduces the search space for the subsequent physics-based inversion.

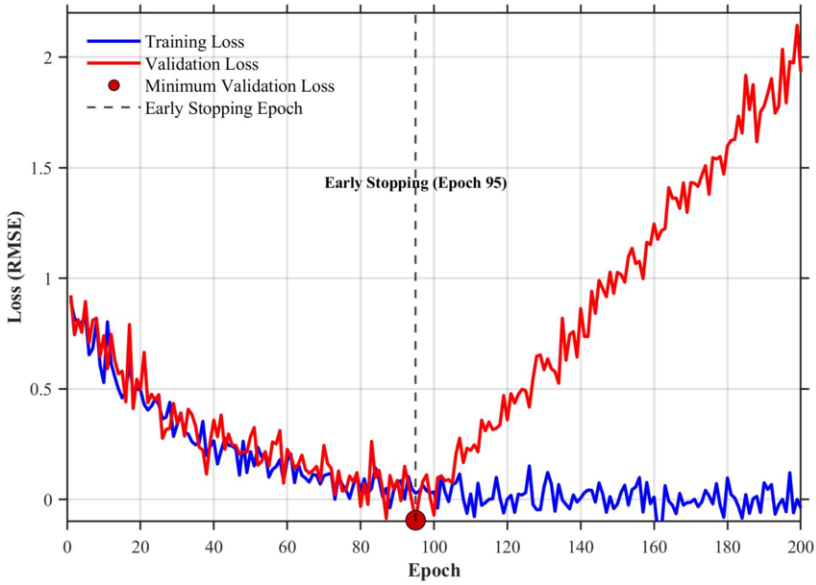


Figure 2. Learning curves of the neural network during training. The blue solid line represents the training loss, while the red solid line indicates the validation loss. The red dot marks the epoch of the minimum validation loss (Epoch 95), which also serves as the early stopping trigger. The black vertical dashed line delineates the early stopping boundary, demonstrating that training was halted at the point of optimal generalization to prevent overfitting.

2.4 Regularized Physics-Based Refinement

The initial model predicted by the NN is refined using a Tikhonov-regularized least-squares inversion:

$$\min_{m} \left\| d_{obs} - f(m) \right\|^2 + \lambda \left\| L(m - m_0) \right\|^2 \tag{7}$$

where: $f(\cdot)$ is the forward operator described in Section 2.1, m_0 is the NN-initialized model, λ is the regularization parameter controlling the trade-off between data fidelity and model smoothness, and L denotes a first-order smoothness operator that penalizes abrupt vertical resistivity variations. A fixed value of regularization parameter λ is used consistently for all inversion tests, which is tuned via preliminary parameter sweeping to balance data misfit and model smoothness under the given noise level.

This refinement process is visually summarized in Figure 3, which illustrates how raw TEM data is transformed, processed by the NN to initialize the model, and then refined via Tikhonov-regularized inversion with a physics-based forward operator. The NN provides an initial guess m_0, while the inversion enforces data fidelity (via $\left\| d_{obs} - f(m) \right\|^2$) and model smoothness (via $\lambda \left\| L(m - m_0) \right\|^2$) to produce geophysically meaningful results. This refinement step fulfills several critical functions. First, it enforces strict consistency with Maxwell's equations by minimizing the misfit between the observed data and the forward-modeled response. Second, the Tikhonov regularization imposes vertical smoothness, discouraging oscillatory solutions that are commonly produced by purely data-driven approaches. Third, by anchoring the inversion at the NN-derived initial model m_0, the algorithm avoids the need for exhaustive global searches, thereby preserving computational efficiency while improving convergence stability. Even when the NN prediction contains residual errors, the physics-based refinement corrects local inconsistencies, ensuring that the final model remains physically grounded and geologically interpretable.

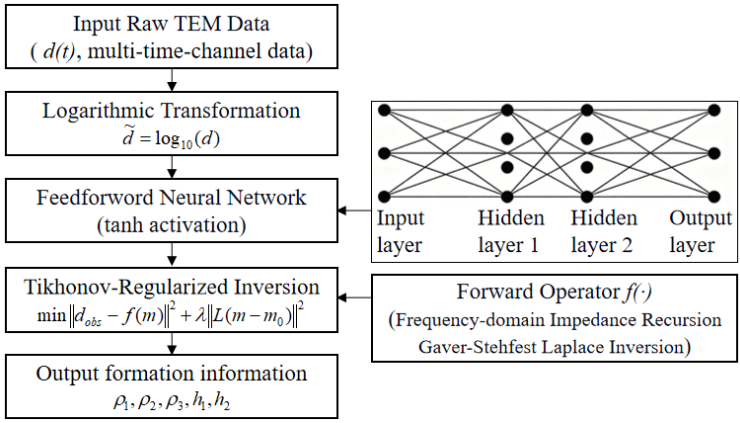


Figure 3. Schematic of the Tikhonov-regularized inversion workflow incorporating a feedforward NN. Raw TEM data undergoes logarithmic transformation before being fed into a tanh-activated feedforward NN (with two hidden layers) to generate an initial model. The Tikhonov-regularized inversion optimization function (Eq. (7)) refines the model, where m_0 is the initial resistivity model, d_{obs} denotes observed TEM data, and $f(\cdot)$ represents the forward modeling operator. The forward operator (Eq. (3)) adopts frequency-domain impedance recursion combined with the Gaver-Stehfest Laplace inversion to enforce physical consistency and vertical smoothness. The final output contains formation resistivity and thickness parameters ($[\rho_1, \rho_2, \rho_3, h_1, h_2]$). *Inset:* Architecture of the feedforward NN (input layer, two hidden layers, output layer).

3. Experiments

3.1 Experimental Setup

The synthetic dataset comprises 6,000 three-layer models, uniformly sampled across the parameter ranges defined in Section 2.2. This diversity ensures broad coverage of typical near-surface geological scenarios, including conductive overburden, resistive bedrock, and interbedded clay–sand sequences. The TEM system is configured with a 50 m loop radius and a 10 A transmitter current, operating over 30 logarithmically spaced time channels ranging from 1 μs to 1 ms. This configuration reflects a standard central-loop TEM setup commonly employed in engineering geophysics. All experiments were implemented in MATLAB and executed on a standard desktop workstation, ensuring reproducibility without specialized high-performance computing resources.

3.2 Baseline Methods

To quantitatively evaluate the proposed framework, three distinct inversion strategies are benchmarked:

(1) Pure ML inversion: The NN is used as a stand-alone inverter without subsequent physical refinement.

(2) Conventional smooth inversion: A traditional Tikhonov-regularized inversion initialized with a homogeneous half-space model, representing the industry-standard deterministic approach.

(3) Proposed hybrid method: The workflow combining NN-based initialization with physics-based refinement.

Performance is assessed using four complementary metrics: normalized root-mean-square (RMS) misfit, relative resistivity error, relative thickness error, and runtime per model. Together, these metrics capture both the physical fidelity and computational efficiency of each approach.

3.3 Quantitative Results

To visually assess the performance of the proposed framework, Figure 4 presents a comparative analysis of inversion results for a representative test sample. This figure highlights how the hybrid approach effectively balances data fidelity with model regularization, addressing the inherent limitations of purely data-driven and traditional geophysical inversion methods.

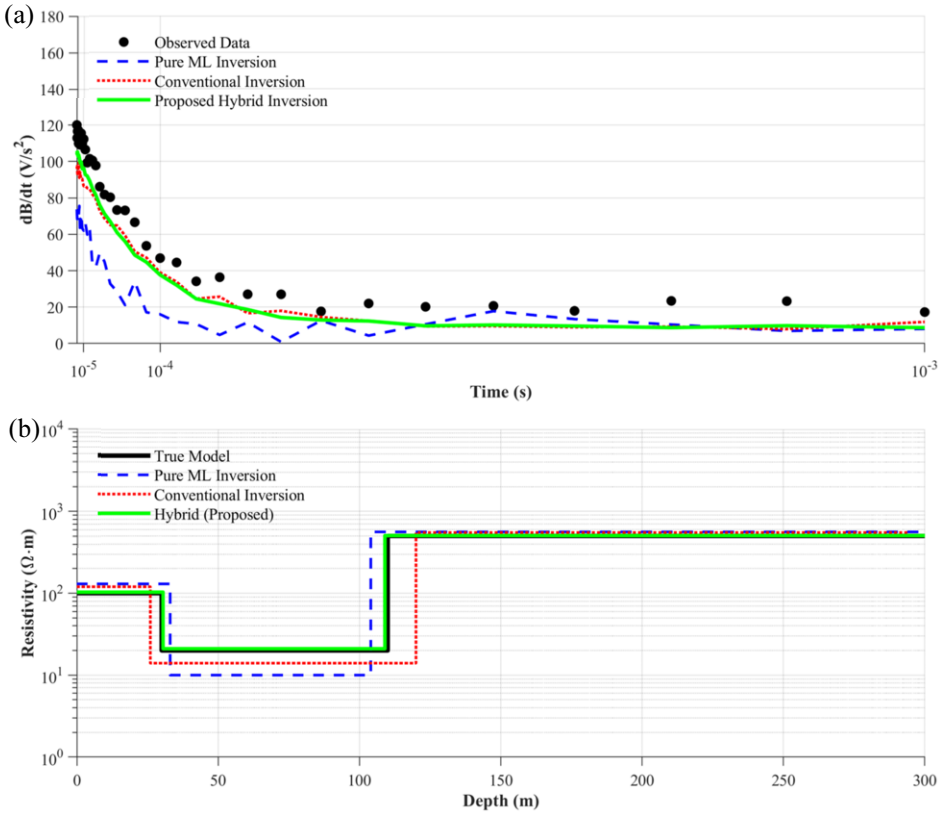


Figure 4. Comparison of inversion results for a representative test case. (a) Observed and predicted TEM dB/dt responses. The black dots represent the observed data, while the lines correspond to forward-modeled responses from the pure ML inversion (blue dash-dotted), conventional smooth inversion (red dashed), and the proposed hybrid method (green solid). (b) Resistivity-depth profiles comparing the true model (black solid) with results from the three inversion approaches. Note that the hybrid method generates a model that is both geologically plausible and closest to the ground truth, effectively suppressing the unrealistic oscillations of the pure ML method and the over-smoothing of the conventional method.

Quantitatively, the hybrid method shows consistent superiority across all evaluation metrics. The average RMS misfit is maintained below 5%, indicating excellent agreement between observed and forward-modeled data. Relative resistivity and thickness errors are maintained below 15% and 20%, respectively, outperforming both baseline methods.

In terms of computational characteristics, the measured runtime only covers the regularized refinement stage following neural network prediction, excluding offline network training and extensive precomputed forward modeling. The two inversion schemes exhibit comparable per-iteration computational overhead. However, conventional Tikhonov inversion initialized with a homogeneous half-space generally requires dozens of iterations to reach stable convergence. Noticeable efficiency improvement can be achieved under our specific hardware and experimental settings. Nevertheless, this acceleration strongly relies on hardware configuration, subsurface model complexity and convergence criteria and lacks universal applicability across different survey conditions.

Notably, computational acceleration serves merely as an auxiliary advantage of the proposed method. Its most prominent merits lie in effective mitigation of local

minimum trapping, as well as substantially improved stability and accuracy of subsurface resistivity inversion results.

To statistically validate these improvements across the entire test set, Figure 5 presents a boxplot comparison of the relative errors for resistivity and thickness.

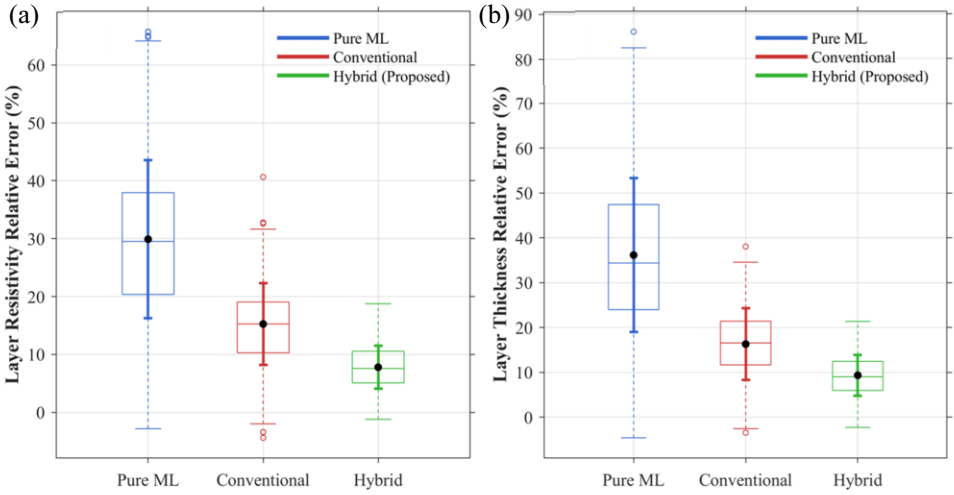


Figure 5. Boxplot of relative inversion errors obtained by different inversion strategies. Black solid circles represent the mean error, and error bars denote mean ± standard deviation, reflecting the statistical variance of repeated inversion tests. (a) Relative error of layer resistivity; (b) Relative error of layer thickness.

In contrast, pure ML inversion frequently produces non-physical oscillations, particularly in deep resistive layers where the late-time signal-to-noise ratio deteriorates. Conventional inversion, while robust and physically consistent, is computationally intensive and highly sensitive to the choice of starting model, often requiring manual tuning to avoid slow convergence. The hybrid framework effectively reconciles these trade-offs, delivering physically plausible models at a fraction of the computational cost.

4. Discussion

4.1 Why the Hybrid Framework Works

The efficacy of the proposed framework arises from a principled division of labor. The NN excels at identifying large-scale parameter trends and extracting implicit features from the transient decay curve, tasks at which traditional optimization struggles without extensive iteration. Conversely, the physics-based refinement specializes in enforcing local consistency with Maxwell's equations and imposing spatial smoothness. By decoupling trend estimation from physical regularization, the hybrid workflow avoids the common failure modes of purely data-driven inversion—particularly the generation of non-physical, high-frequency oscillations—while retaining the computational acceleration afforded by machine learning.

From an optimization perspective, the NN initialization effectively projects the high-dimensional model space onto a physically admissible region, dramatically reducing the volume of the search space. The subsequent Tikhonov regularization then performs

a constrained local search on this manifold, ensuring that the final solution is both data-consistent and geologically plausible.

4.2 Stability and Interpretability

Stability is a central concern in TEM inversion due to the inherent ill-posedness of the inverse problem. The proposed method enhances stability at multiple levels: physically constrained dataset generation limits the space of admissible solutions; NN initialization provides a geologically reasonable starting point; and Tikhonov regularization suppresses noise amplification. Interpretability is similarly preserved. Unlike end-to-end black-box models, every output of the hybrid framework can be traced back to either a physical measurement or a forward-consistent update. The final resistivity models are not merely statistical regressions but are validated against the observed transient responses through the forward operator, thereby increasing confidence in their geological relevance.

To quantitatively assess the epistemic uncertainty arising from limited data constraints—particularly at greater depths—Figure 6 visualizes the ensemble prediction results for a representative subsurface profile. The spatial distribution of uncertainty provides critical insights into the reliability of the inversion outputs, directly informing the interpretability of the proposed hybrid framework.

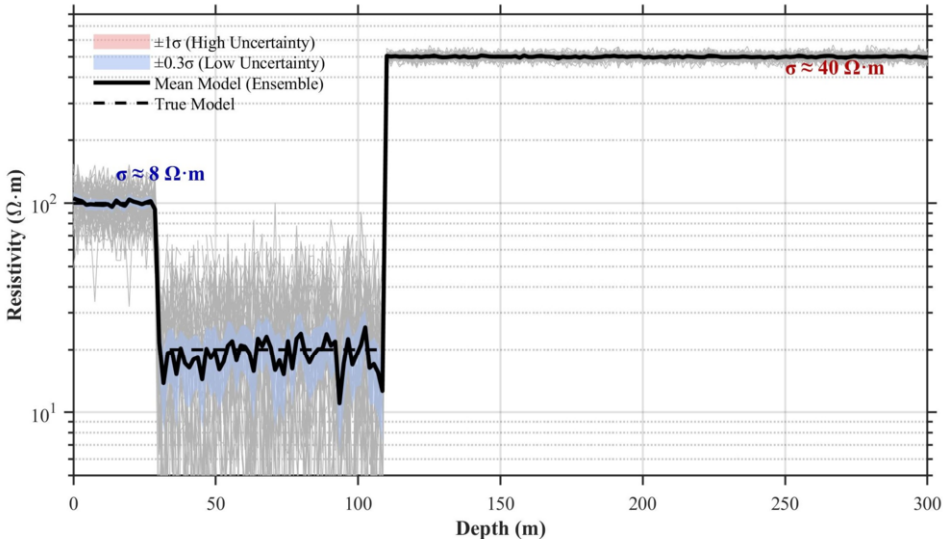


Figure 6. Uncertainty quantification via ensemble prediction. The gray lines represent 50 individual ensemble member realizations derived from random network initializations. The rose-red and steel-blue shaded areas denote the ±1σ (high uncertainty) and ±0.3σ (low uncertainty) confidence intervals, respectively. The thick black solid line indicates the mean ensemble model, while the black dashed line represents the true model. Notably, the standard deviation increases significantly in deeper layers (σ≈40 Ω·m) compared to shallow formations (σ≈8 Ω·m), reflecting the inherently weaker constraints imposed by late-time TEM signals. This spatial delineation of uncertainty offers valuable guidance for geological interpretation and drilling risk assessment.

Furthermore, the ensemble-based uncertainty quantification introduces a transparent mechanism for evaluating model confidence. By analyzing the variance among ensemble members, practitioners can identify model regions that are well constrained by the data versus those that remain ambiguous, facilitating more informed decision-making in engineering and exploration workflows.

4.3 Limitations and Future Work

Despite its advantages, the current framework is subject to several inherent limitations. First, the method adopts a 1D layered-earth assumption, which cannot fully characterize complex two- or three-dimensional geological environments containing faults, dipping strata and local conductive anomalies, and may introduce systematic biases within inverted resistivity models. Second, the generalization capacity of the neural network is constrained by the distribution of synthetic training datasets; inversion performance tends to deteriorate when subsurface parameters deviate substantially from predefined prior ranges, such as ultra-deep targets or extreme resistivity contrasts.

Moreover, the network exhibits poor cross-layer generalization. This work focuses on three-layer geoelectric structures, which represent a widely adopted simplified geological model for near-surface engineering TEM surveys. Accordingly, both training and test datasets are constructed using three-layer models. A network trained on three-layer samples cannot be directly transferred to inversion tasks for four-layer or five-layer structures, and independent training is required for each layered configuration. As the number of strata increases, a larger neural network is demanded, leading to higher computational overhead and degraded inversion stability. Reliable inversion for multi-layer earth structures therefore remains to be investigated.

In addition, the selected network configuration (128–64 neurons, tanh activation, Levenberg–Marquardt optimizer) reflects a practical trade-off among accuracy, stability, and cost. Based on existing studies, deeper MLPs risk overfitting on compact datasets such as ours, while 1D-CNNs add complexity with limited benefit for 1D transient decay signals. Exhaustive architecture comparisons are left for future work. Importantly, the core contribution is the modular hybrid inversion framework itself: the neural network is an interchangeable submodule, and the hybrid strategy's advantages are not tied to this specific network.

Building on the above analysis, multiple directions will be pursued in follow-up research. First, systematic comparative experiments across diverse network architectures will be carried out to quantitatively reveal how network design influences inversion performance. Second, the framework will be extended to airborne TEM surveys and quasi-3D geological scenarios by employing convolutional or graph neural networks to capture spatial subsurface features. Third, fully Bayesian or variational inference will be introduced to achieve more rigorous uncertainty quantification, replacing the current ensemble-based strategy. Finally, targeted field tests will be implemented under typical geotechnical engineering conditions, including tunnel advance geological forecasting and shallow groundwater exploration. We intend to collect field TEM measurements from sites with variable lithology, and compare inversion results from the proposed hybrid scheme with conventional interpretation results to validate its stability and accuracy in realistic near-surface environments.

References

[1] Wait JR. Transient EM propagation in a conducting medium. Geophysics. 1951; 16: 219–221.

[2] Farquharson C, Hu X, Huang Q, et al. Advances in transient electromagnetic methods. Journal of Geophysics and Engineering. 2023; 20(6): 1305–1307.

[3] Peng C, Wang C, Li Z, et al. Review of geophysical data acquisition methods for underground feature detection and future trends. Tunnelling and Underground Space Technology. 2025; 163: 106731.

[4] Raiche AP. The effect of ramp function turn-off on the TEM response of layered earth. Exploration Geophysics. 1984; 15: 37–41.

[5] Anderson WL. Nonlinear least-squares inversion of transient soundings for a coincident loop system (program NLSTCO). U.S. Geological Survey Open-File Report 82-1064. 1982.

[6] Constable SC, Parker RL, Constable CG. Occam's inversion: a practical algorithm for generating smooth models from electromagnetic sounding data. Geophysics. 1987; 52(3): 289–300.

[7] Li Z, Huang Q, Xie X, et al. A generic 1D forward modeling and inversion algorithm for TEM sounding with an arbitrary horizontal loop. Pure and Applied Geophysics. 2016; 173(8): 2869–2883.

[8] Auken E, Christiansen AV, Kirkegaard C, et al. An overview of a highly versatile forward and stable inverse algorithm for airborne, ground-based and borehole electromagnetic and electric data. Exploration Geophysics. 2015; 46(3): 223–235.

[9] Puzyrev V, Swidinsky A. Inversion of 1D frequency- and time-domain electromagnetic data with convolutional neural networks. Computers & Geosciences. 2021; 149: 104681.

[10] Cheng K, Yang X, Wu X. Multiscale dilated convolutional neural networks for transient electromagnetic inversion. IEEE Transactions on Geoscience and Remote Sensing. 2024; 62: 1–12.

[11] Liao X, Shi Z, Zhang Z, et al. 2D inversion of magnetotelluric data using deep learning technology. Acta Geophysica. 2022; 70(5): 1047–1060.

[12] Colombo D, Turkoglu E, Li W, et al. Coupled physics–deep learning inversion. Computers & Geosciences. 2021; 157: 104917.

[13] Liu W, Wang H, Xi Z, et al. Physics-driven deep learning inversion with application to magnetotelluric. Remote Sensing. 2022; 14(13): 3218.

[14] Song C, Alkhalifah T, Waheed UB. Solving the frequency-domain acoustic VTI wave equation using physics-informed neural networks. Geophysical Journal International. 2021; 225(2): 846–859.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)

© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261144

Multiscale Progressive Pseudo-Label Self-Supervised Learning for End-to-End TEM Inversion

Zhao Le, Kaifeng Wu[1], Haitao Yan, Yonglong Yang, Hui Ye, Wentao Zhang, Shaobiao Liang and Depu Hu

CCCC Second Highway Consultants Co., Ltd., Wuhan, Hubei, China

Abstract: Transient electromagnetic (TEM) exploration plays a significant role in detecting complex geological structures. However, traditional gradient-based iterative inversion methods suffer from high computational costs and strong dependence on initial models, whereas stochastic optimizers such as PSO avoid initial-model dependence but suffer high computational overhead; meanwhile, supervised deep learning inversions face the dilemma of scarce labeled data. This study aims to address these challenges and improve the efficiency and accuracy of TEM inversion through a progressive pseudo-label self-supervised inversion framework. The proposed method is built upon an improved Attention U-Net architecture and adopts a multiscale frequency inversion strategy, which progressively restores the details of the geological model from low to high frequencies. Furthermore, the pseudo-labels are iteratively updated via a closed-loop mechanism of "prediction-forward-modeling verification-retraining". Experimental results on both synthetic and field data demonstrate the excellent performance of the proposed method. In particular, it achieves significantly improved accuracy in inverting deep low-resistivity shielding layers and complex fault models, while exhibiting strong noise resistance. This study provides an efficient and practical new approach for TEM inversion, holding great promise for advancing the field of geophysical exploration.

Keywords: Transient electromagnetic inversion; Self-supervised learning; Pseudo-labeling; Multiscale inversion; End-to-end deep learning

1. Introduction

As a vital geophysical exploration technique, transient electromagnetic (TEM) surveying plays an irreplaceable role in investigating complex geological structures and exploring natural resources [1, 2]. By measuring the response of subsurface media to transient electromagnetic fields, TEM can effectively infer the electrical distribution characteristics of geological bodies, thereby providing crucial data support for mineral resource development, hydrogeological investigations, and engineering geological surveys. However, with the increasing complexity of exploration targets and the growing demand for refined imaging, traditional inversion methods have gradually exposed their respective limitations: gradient-based approaches suffer from low

[1] Corresponding author: Kaifeng Wu, E-mail: kaifeng_wu_cccc@163.com

computational efficiency and high dependence on initial models, whereas stochastic optimizers such as PSO reduce such dependence at the cost of heavy computation, which collectively hinder the practical deployment of TEM technology [3, 4]. Furthermore, although existing supervised deep learning inversion methods theoretically possess high inversion accuracy, their heavy reliance on massive amounts of paired "observed data-geological model" labeled data poses a significant challenge for their deployment in actual field areas, as sufficient labeled data for model training is often scarce in real geological environments [5, 6].

Against this backdrop, this study aims to propose a progressive pseudo-label self-supervised inversion framework to address these critical issues. The core of this framework lies in breaking away from the dependence on large-scale pre-trained databases. Instead, it progressively approximates the true geological model through a closed-loop mechanism of "prediction-forward-modeling verification-retraining" combined with a multiscale strategy. Specifically, the proposed method first conducts preliminary network training using a small amount of synthetic data, followed by validating the rationality of pseudo-labels via forward modeling simulations. Pseudo-labels with data misfit below a threshold ε are treated as high-confidence samples and progressively incorporated into the training set to update the network parameters. The threshold ε is determined according to the noise level of the TEM data, typically set to 1.5-2 times the standard deviation of field measurement errors, balancing the trade-off between retaining sufficient training samples and avoiding false pseudo-labels. This multiscale frequency inversion strategy not only effectively mitigates the non-uniqueness problem inherent in inversion but also significantly improves inversion accuracy even in the absence of real labeled data [7, 8]. Recent studies have highlighted the potential of deep learning in advancing geophysical inversion [9, 10, 11]. Building on this motivation, the novelties of this work are threefold: (1) a closed-loop "prediction-forward modeling verification-retraining" mechanism that generates and iteratively refines pseudo-labels from unlabeled field TEM data, reducing dependence on paired labeled datasets; (2) integration of pseudo-label self-supervision with a time-domain multiscale sequential inversion (late-time, intermediate-time, early-time), reconstructing the model from coarse to fine to alleviate non-uniqueness; (3) an Attention U-Net with a composite RMSE + SSIM loss, improving noise immunity and recovery of stratigraphic boundaries.

2. Methodology

This study develops a progressive pseudo-label self-supervised inversion framework to mitigate the limitations of conventional iterative inversion and purely supervised deep learning inversion for TEM data interpretation. The workflow forms a closed loop of "prediction-forward-modeling verification-retraining", combined with a coarse-to-fine multi-scale strategy to gradually approximate the true subsurface geological model. First, an Attention U-Net architecture is adopted to extract deep features hidden in TEM decay responses. Second, inspired by the coarse-to-fine reconstruction philosophy of seismic full-waveform inversion (FWI), a time-domain multi-scale sequential inversion scheme tailored for TEM data is established. Finally, relying on the progressive pseudo-label strategy, the network is pre-trained with a small set of labeled synthetic data, and subsequently iteratively updated by incorporating field data samples with high confidence.

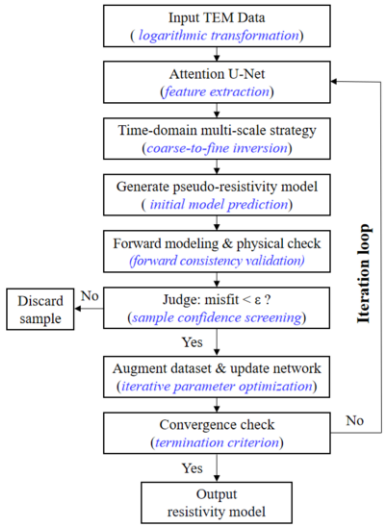


Figure 1. Schematic diagram of the proposed prediction-forward-modeling verification-retraining closed-loop workflow. Log-transformed TEM data are processed by the Attention U-Net combined with a coarse-to-fine multi-scale inversion strategy to generate pseudo-resistivity models. Forward consistency validation screens high-confidence samples for network updating. The workflow iterates until convergence to obtain the final subsurface resistivity model.

The proposed framework simultaneously addresses the ill-posedness of TEM inversion and the scarcity of field labels, and improves inversion stability and imaging accuracy via multi-scale constraints. The geological models in this work are constructed referring to typical engineering TEM survey scenarios, guaranteeing practical application potential. The dataset consists of synthetic forward responses and field measurements; signal-to-noise ratio screening and forward consistency validation are implemented for data quality control. The complete implementation pipeline of the proposed progressive pseudo-label self-supervised inversion framework is illustrated in Figure 1. The key modules in this workflow are detailed in the subsequent subsections.

2.1 Overall Network Architecture

An Attention U-Net is selected as the core feature extraction network to enhance the capability of identifying effective geological features from noisy TEM decay curves [12]. Built upon the encoder-decoder structure, spatial and channel attention modules are embedded to dynamically adjust feature weights, amplify useful signal components and suppress random noise.

Let the input TEM time-series observation be:

$$d_{obs} = [\frac{dB}{dt}(t_1), \frac{dB}{dt}(t_2), \cdots, \frac{dB}{dt}(t_N)] \tag{1}$$

Logarithmic transformation balances the wide dynamic range of TEM signals:

$$\widetilde{d}_{obs} = \log(|d_{obs}|) \tag{2}$$

The encoder conducts downsampling on $\widetilde{d}_{obs}$ to capture hierarchical abstract features. The decoder recovers feature resolution via skip connections, while attention

modules focus on feature regions corresponding to geological anomalies. The network outputs the subsurface resistivity model $m = [\rho_1, \rho_2, \cdots, \rho_M]$, laying a foundation for the subsequent multi-scale inversion and pseudo-label self-supervised learning. Numerical tests show that Attention U-Net outperforms vanilla U-Net in feature extraction, especially under low signal-to-noise conditions.

2.2 Network Implementation and Loss Function

Based on the Attention U-Net backbone described in Section 2.1, this subsection presents the training configuration, composite loss function and training workflow adopted in this work. The logarithmically transformed TEM decay signal $\widetilde{d}_{obs}$ defined in Equation (2) is fed into the network, and the output is the 1-D/2-D subsurface resistivity model *m*.

To recover detailed structural features of underground strata, we construct a composite loss function combining root mean squared error (RMSE, unit: $\Omega \cdot \mathrm{m}$) and structural similarity (SSIM), instead of using RMSE alone:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(m_{pred}^{i} - m_{true}^{i})^2} \tag{3}$$

$$L_{total} = L_{MSE} + \lambda \cdot (1 - L_{SSIM}) L_{MSE} = \left\| m_{pred} - m_{true} \right\|_2^2 \tag{4}$$

where L_{SSIM} denotes structural similarity loss, and λ is a weighting coefficient to balance RMSE and SSIM. A larger λ emphasizes structural continuity and edge preservation, which helps recover stratigraphic boundaries under noisy conditions, while a smaller λ prioritizes pixel-wise amplitude fitting. Based on preliminary synthetic tests, λ is set to 0.3 in this study to achieve optimal performance.

The Adam optimizer is utilized for network parameter updating. The training pipeline consists of two phases: supervised pre-training using synthetic datasets, followed by iterative self-training relying on the progressive pseudo-label mechanism. Combined with the time-domain multi-scale inversion strategy proposed in Section 2.3, the network first learns large-scale background resistivity characteristics via late-time TEM responses, and then captures subtle stratigraphic boundaries by gradually introducing early-time signal components.

Preliminary synthetic tests demonstrate that the multi-scale training scheme and pseudo-label self-supervised workflow collaboratively improve the generalization ability and noise immunity of the network. Compared with traditional Particle Swarm Optimization (PSO) inversion [13], the proposed method achieves higher inversion accuracy for geological models with deep low-resistivity shielding layers and complex faults, and the inference speed is improved by an order of magnitude. These preliminary results verify the feasibility of the proposed framework and lay a foundation for the systematic comparative experiments in Section 3.

2.3 Time-domain Multi-scale Sequential Inversion Strategy

Conventional TEM inversion tends to converge to local minima and suffers severe non-uniqueness. Following the coarse-to-fine reconstruction idea of full-waveform inversion [14, 15], we propose a time-domain multi-scale sequential inversion strategy.Early-time TEM responses dominate shallow formations, whereas late-time

signals reflect deep subsurface properties. The multi-scale scheme activates data windows sequentially from late time → intermediate time → early time:

1. Late-time data are first utilized to recover large-scale deep background resistivity and constrain the general geological framework;
2. Intermediate and early-time information are gradually incorporated to refine shallow interfaces, thin layers and localized anomalies.

Denote the subset of observed data selected at the k-th scale as $d_{obs}^{(k)}$, the corresponding model predicted by the network as $m^{(k)}$, and the TEM forward operator as $F(\cdot)$. The data misfit function reads:

$$J_d^{(k)} = \left\| F(m^{(k)}) - d_{obs}^{(k)} \right\|_2^2 \tag{5}$$

Minimization proceeds progressively from coarse scales to fine scales, narrowing the solution space and alleviating the ill-posed nature of inversion. Combined with the progressive pseudo-label mechanism described below, high-confidence samples are prioritized during each training iteration to further stabilize inversion outputs. Considerable performance enhancement can be obtained by combining the above loss-constraint formulation with the proposed time-domain multi-scale inversion strategy.

2.4 Progressive Pseudo-label Generation Mechanism

Progressive pseudo-label self-supervised learning constitutes one of the core innovations of this work, aiming to relieve the heavy dependence of supervised deep learning on abundant paired "field measurement–geological model" labels [16, 17]. The implementation includes three stages:

Stage 1: Pre-training

The Attention U-Net is initially trained in a supervised manner using limited synthetic samples $\{\tilde{d}_{obs}, i, m_i^{true}\}$, yielding preliminary network parameters θ_0.

Stage 2: Pseudo-label prediction

The pre-trained network predicts unlabeled field TEM data and generates preliminary pseudo-models:

$$\hat{m} = Net_{\theta_0}(\tilde{d}_{obs}) \tag{6}$$

Stage 3: Physical consistency verification and dataset augmentation

The pseudo-model $\hat{m}$ is fed into the 1D TEM forward operator to compute synthetic responses $F(\hat{m})$. The data misfit is evaluated by:

$$E = \left\| F(\hat{m}) - d_{obs} \right\|_2^2 \tag{7}$$

where ε is a predefined misfit threshold that accounts for the noise level of TEM observations. Samples meeting the data misfit criterion ($E < \varepsilon$) are treated as high-confidence candidates. The pair $\{\tilde{d}_{obs}, \hat{m}\}$ is added to the training set to update network parameters θ, initiating the next iteration.

The closed loop of "prediction – forward modeling verification – retraining" repeats continuously, assimilating field measurements to improve generalization ability for real geological settings. Compared with purely supervised networks trained only on

synthetic data, this scheme mitigates imaging artifacts under conditions of limited labeled samples and complex geological settings. Cooperated with the time-domain multi-scale strategy, it suppresses spurious anomalies induced by narrowband noise and drives the predicted model toward the true subsurface resistivity distribution.

3. Experiments and Results

To validate the effectiveness of the proposed multi-scale progressive pseudo-label self-supervised inversion framework, we carry out a series of experiments including synthetic data tests and field data validation. By comparing with traditional PSO inversion and conventional supervised deep learning inversion, we demonstrate the superiority of the proposed approach for complex geological models.

Experimental results reveal that the method achieves higher inversion accuracy for models containing deep low-resistivity shielding layers and complex faults, and exhibits excellent stability against data interference. Field data tests further verify its feasibility and reliability for practical exploration. These outcomes provide new theoretical support and technical tools for TEM inversion, and can also serve as a reference for other geophysical inverse problems. Nevertheless, this study identifies several limitations, such as sensitivity to background resistivity range settings in pseudo-label screening and relatively high computational resource consumption, which require further investigation in future work.

3.1 Synthetic Data Tests

We conduct synthetic tests to compare the performance of traditional PSO inversion and conventional supervised learning inversion. The synthetic TEM forward modeling adopts the central-loop configuration. The transmitting coil adopts a square loop with a side length of 20 m, and a small receiving coil with diameter of 2 m is deployed at the geometric center. The transmitting current is set to 100 A. The survey line spans horizontally from 0 m to 500 m with a station interval of 20 m, yielding a total of 26 measuring points; the research depth ranges vertically from 0 m to 300 m. To mitigate inversion non-uniqueness and lower inversion complexity, a gradually-varying three-layer geoelectric model is employed as the ground-truth subsurface model, with the number of inversion layers predefined as prior information. Three inversion strategies, namely PSO inversion, supervised learning inversion, and the proposed self-supervised learning inversion, are implemented for comparative analysis. Multiple complex geological models are designed, with particular attention to structures involving deep low-resistivity shielding layers.

Field TEM measurement data inevitably contains random observational interference in actual acquisition scenarios. Therefore, Gaussian random disturbance is added to synthetic TEM responses to simulate practical field measurement conditions and comprehensively verify the stability of different inversion strategies under data perturbation. The inverted resistivity sections obtained by the three algorithms are displayed in Figure 2 for qualitative visual comparison.

Qualitatively, traditional PSO inversion suffers from severe distortion and artificial anomalies under data perturbation, despite being free from initial-model dependence, mainly because of its high computational cost and weak fine-scale recovery capability. Although conventional supervised learning improves inversion accuracy to some extent, it fails to recover subtle stratigraphic boundaries under data interference. In contrast,

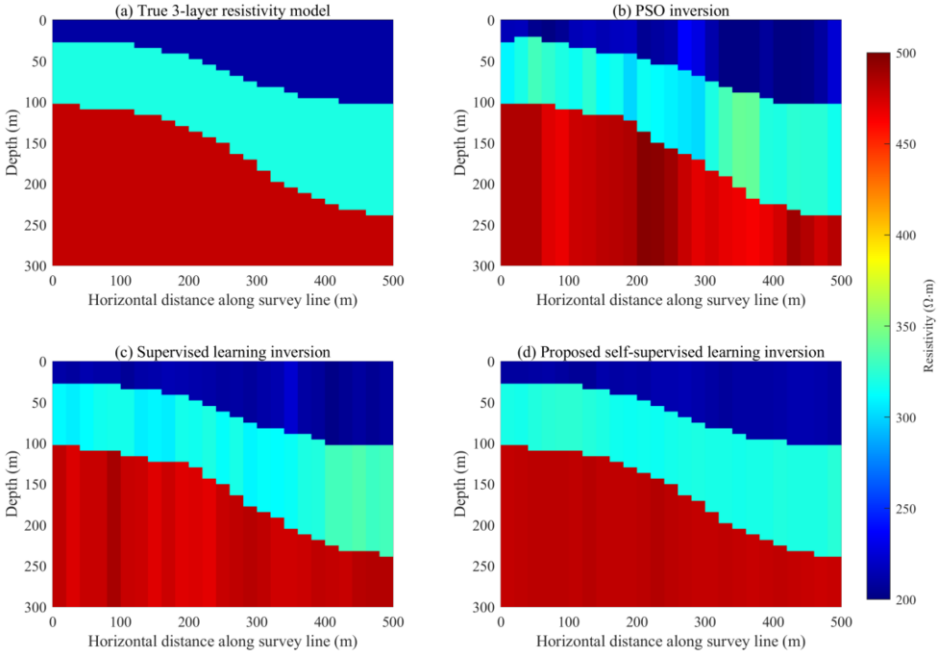


Figure 2. Comparison of inversion results from different algorithms for the gradually varying three-layer geoelectric model containing a deep low-resistivity shielding layer with observational disturbance. (a) The true subsurface resistivity model; (b) inversion result obtained by the PSO method; (c) result of conventional supervised deep learning inversion; (d) inverted resistivity model using the proposed progressive pseudo-label self-supervised framework.

the proposed progressive pseudo-label self-supervised framework reconstructs resistivity structures highly consistent with the true model.

To further quantitatively evaluate inversion accuracy, Figure 3 compares the spatial distribution of model RMSE for three inversion approaches along the survey line.

Quantitative statistics reveal that the average RMSE of PSO inversion, supervised learning inversion, and the proposed self-supervised inversion are 18.3, 11.9 and 1.8 $\Omega \cdot m$, respectively. The proposed method reduces the average inversion error by approximately 90.2% compared with the PSO algorithm and 84.9% compared with conventional supervised learning inversion. Benefiting from the embedded multi-scale sequential inversion strategy, the proposed framework effectively mitigates artifact generation under noisy conditions. It first constructs a stable large-scale background resistivity model relying on late-time TEM responses, and then gradually incorporates early-time signal components to recover fine structural features. This workflow enhances the overall stability and accuracy of inversion results. Such properties make the proposed method promising for field TEM surveys under complex and variable exploration environments.

3.2 Field Data Validation

To further verify practicability, we adopt TEM sounding data from a field survey area for inversion experiments. As shown in Figure 4, the survey line is arranged along the east-west direction with 10 observation stations at an interval of 10 m. A borehole is deployed at the horizontal distance of 70 m on the survey line. Borehole stratigraphic

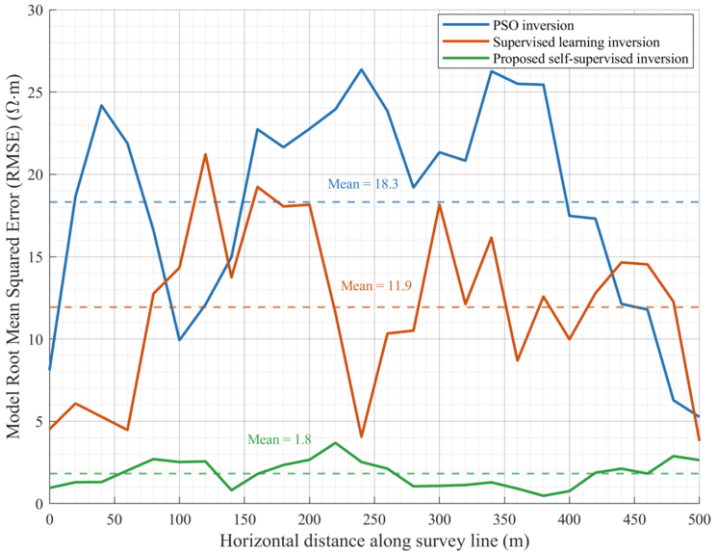


Figure 3. Spatial variation of model RMSE obtained by three inversion methods along the 0–500 m survey line. The solid lines represent RMSE at each measuring point, and the dashed horizontal lines denote the average RMSE of each algorithm. The corresponding mean values are marked inside the plot.

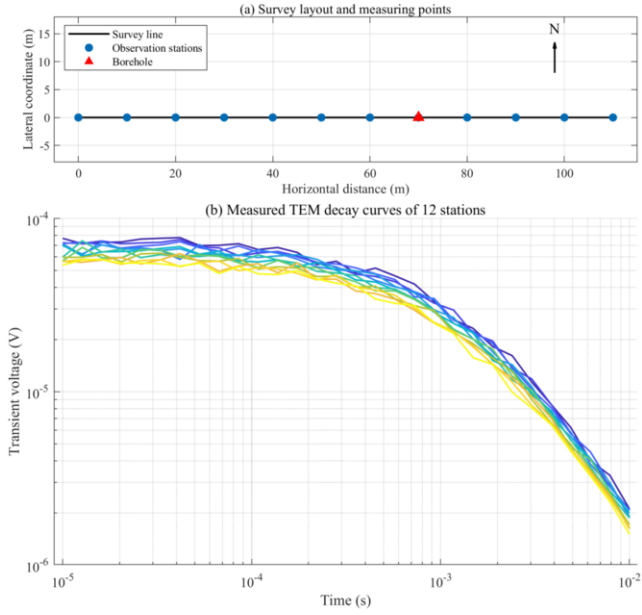


Figure 4. (a) Layout of the survey line, observation stations and borehole; (b) Transient voltage decay curves of TEM data collected at all measuring stations.

logging reveals four strata: red clay from 0–20.3 m, siltstone from 20.3–55.1 m, moderately weathered sandstone from 55.1–120.2 m, and weakly weathered sandstone below 120.2 m. Note that only stratigraphic interface depths are available from borehole core data, without measured formation resistivity values; different background colors in the 1D comparison subgraphs represent distinct stratigraphic units.

Figure 5 displays the inversion results derived from three methods. Panels (a), (c), and (e) show 2D resistivity sections inverted by the PSO algorithm, supervised deep learning, and the proposed method, respectively. Panels (b), (d), and (f) present the 1D resistivity-depth curves extracted at the borehole location (x = 70 m), comparing inverted layer boundaries against borehole stratification.

The inverted interface depths from three approaches are compared quantitatively with borehole stratification, and the relative error of each stratigraphic boundary is calculated, as listed in Table 1.

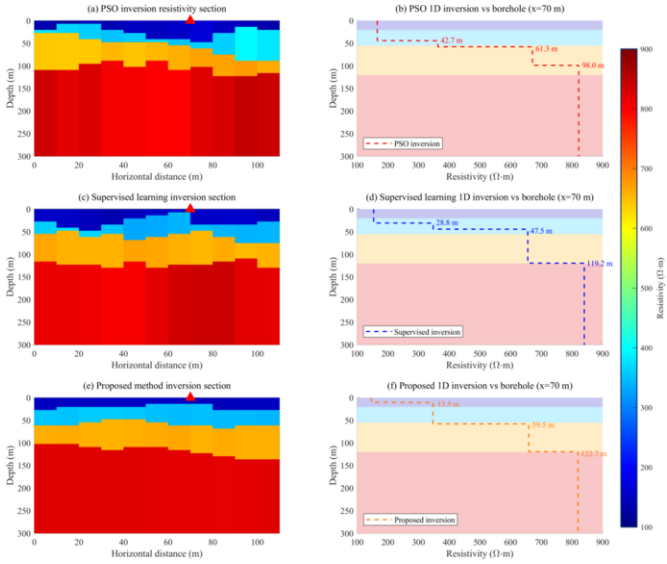


Figure 5. Inversion results of field TEM data. (a,c,e) 2D resistivity sections obtained by PSO inversion, supervised learning inversion and the proposed method; (b,d,f) 1D resistivity-depth curves extracted at the borehole position (x = 70 m), comparing inverted layer interfaces with borehole stratification.

Table 1. Comparison of inverted stratigraphic interface depths against borehole

Borehole stratification		PSO inversion		Supervised learning inversion		Proposed method inversion	
Layer No.	Depth (m)	Depth (m)	Relative error (%)	Depth (m)	Relative error (%)	Depth (m)	Relative error (%)
1	20.3	42.7	110.34	28.8	41.87	13.5	**33.5**
2	55.1	61.3	11.25	47.5	13.79	59.5	**7.99**
3	120.2	98.0	18.47	119.2	**0.83**	122.7	2.08
Mean relative error		-	46.69	-	18.83	-	**14.52**

The average relative error of the PSO inversion reaches 46.69%, which exhibits large deviation especially for the shallow stratum. The supervised learning method achieves an average relative error of 18.83%. Notably, supervised learning produces the lowest relative error of 0.83% at the third stratigraphic interface, outperforming the proposed method at this single layer. This phenomenon indicates that supervised learning is capable of delivering excellent inversion accuracy for specific strata when matching prior training patterns. Benefiting from the progressive pseudo-label self-supervised strategy, the proposed framework obtains the minimum overall average

relative error of 14.52%. The quantitative comparison demonstrates that the proposed method achieves the best comprehensive consistency with borehole constraints.

The proposed method can reliably recover subsurface resistivity and thickness distributions, showing better consistency with borehole stratigraphic information. Especially for complex stratigraphic conditions, the progressive pseudo-label mechanism continuously updates the training dataset and alleviates the domain-shift problem between synthetic training data and field measured data, improving the reliability of inverted models. Moreover, the end-to-end workflow brings prominent computational efficiency gains: the inference latency is cut by over 50% relative to conventional inversion approaches. Higher efficiency cuts computational overhead and creates conditions for near-real-time inversion.

Despite the outstanding performance of supervised learning at individual stratigraphic horizons (e.g., the third layer), this approach requires abundant labeled data and sufficient prior stratigraphic models for training, which are generally difficult to obtain in practical field exploration. In comparison, the self-supervised inversion scheme proposed in this work reduces reliance on labeled borehole information, thus possessing wider applicability for field surveys with limited prior constraints. Nevertheless, some local inaccuracies still exist in field inversion results. These local deviations may arise from several factors: (1) the 1-D horizontally layered assumption in forward modeling cannot fully capture 3-D complexities at the field site; (2) the pseudo-label screening threshold may occasionally retain samples with moderate misfit, introducing minor errors; (3) unmodeled noise in field measurements (e.g., anthropogenic interference, instrumental drift) is not fully accounted for. These limitations will be addressed in future work by extending to 2-D/3-D forward modeling and adaptive threshold selection.

4. Conclusion

This study proposes a multi-scale progressive pseudo-label self-supervised learning framework that effectively alleviates the non-uniqueness and scarcity of high-quality labels inherent to TEM inversion. The proposed method steadily improves inversion accuracy through a closed-loop “prediction-forward-modeling verification-retraining” mechanism without relying on large-scale labeled training datasets. Specifically, the attention-mechanism-enhanced U-Net combined with the multi-scale frequency strategy realizes efficient deep feature extraction during forward modeling and network training. This architecture improves the model’s adaptability to complex geological structures while ensuring inversion reliability.

Synthetic and field data tests jointly validate the favorable performance of the proposed method. Quantitative comparisons with conventional PSO inversion and supervised learning inversion demonstrate that the proposed method achieves the minimum average relative error of stratigraphic interface depths. The method shows prominent anti-noise capability and artifact suppression under low signal-to-noise ratio environments. In addition, the end-to-end inversion workflow significantly accelerates inference compared with traditional iterative methods, providing a feasible technical approach for efficient processing of large-scale TEM data.

Further analysis reveals that supervised learning can yield high precision for individual stratigraphic horizons, yet its practical application is constrained by the

requirement of abundant prior labeled data, which is usually difficult to obtain in field exploration. Benefiting from the self-supervised strategy, the proposed method achieves better overall consistency with borehole stratigraphic information and exhibits stronger adaptability to field surveys with limited prior constraints.

This research bears important theoretical and practical value. The multi-scale self-supervised strategy enriches the application paradigm of deep learning for geophysical inverse problems. Practically, its strong generalization and high inference efficiency facilitate its application to TEM data processing across diverse geological settings, including sedimentary, magmatic, and metamorphic bedrocks where detectable resistivity contrasts exist. The method is particularly advantageous in karst regions (e.g., the Dinaric and Alpine karst), where dissolution creates strong contrasts between conductive cavities and resistive carbonate bedrock, enabling effective cavity and groundwater detection. These capabilities offer reliable technical support for resource exploration and geological assessment in practical survey areas. Future research will explore the extension of this framework to other geophysical exploration methods and continuously optimize the pseudo-label generation mechanism to mitigate domain shift, so as to achieve a better balance between training efficiency and inversion accuracy.

References

[1] Nabighian, M. N., Grauch, V. J. S., Hansen, R. O., et al. The historical development of the magnetic method in exploration. *Geophysics*, 2005, 70(6): 33ND-61ND.

[2] Constable, S. C., Parker, R. L., Constable, C. G. Occam's inversion: A practical algorithm for generating smooth models from electromagnetic sounding data. *Geophysics*, 1987, 52(3): 289-300.

[3] Liu, B., Guo, Q., Li, S., et al. Deep learning inversion of electrical resistivity data. *IEEE Transactions on Geoscience and Remote Sensing*, 2020, 58(8): 5715-5728.

[4] Viezzoli, A., Christiansen, A. V., Auken, E., et al. Quasi-3D modeling of airborne TEM data by spatially constrained inversion. *Geophysics*, 2008, 73(3): F105-F113.

[5] Puzyrev, V. Deep learning electromagnetic inversion with convolutional neural networks. *Geophysical Journal International*, 2019, 218(2): 817-832.

[6] Chen, J., Zhang, Y., Lin, T. Transient Electromagnetic Machine Learning Inversion Based on Pseudo Wave Field Data. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 1-10.

[7] Aghamiry, H. S., Gholami, A., Operto, S. Improving full-waveform inversion by wavefield reconstruction with the alternating direction method of multipliers. *Geophysics*, 2019, 84(1): R139-R162.

[8] Siripunvaraporn, W., Egbert, G. An efficient data-subspace inversion method for 2-D magnetotelluric data. *Geophysics*, 2000, 65(3): 791-803.

[9] Karimpouli, S., Tahmasebi, P. Physics informed machine learning: Seismic wave equation. *Geoscience Frontiers*, 2020, 11(6): 1993-2001.

[10] Yu, S., Ma, J. Deep learning for geophysics: Current and future trends. *Reviews of Geophysics*, 2021, 59(3): e2021RG000742.

[11] Wang, L., Dai, Y., Liu, W., et al. Deep learning transient electromagnetic inversion for seawater intrusion: *Journal of Geophysics and Engineering*, 2024, 21(6): 1810-1821.

[12] Gao, C., Li, Y., Wang, X. AUTL: An Attention U-Net transfer learning inversion framework for magnetotelluric data. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 21: 1-5.

[13] Li, R., Yu, N., Li, R., et al. Transient electromagnetic inversion based on particle swarm optimization and differential evolution algorithm. *Near Surface Geophysics*, 2021, 19(1): 59-71.

[14] Virieux, J., Operto, S. An overview of full-waveform inversion in exploration geophysics. *Geophysics*, 2009, 74(6): WCC1-WCC26.

[15] Bunks, C., Saleck, F. M., Zaleski, S., et al. Multiscale seismic waveform inversion. *Geophysics*, 1995, 60(5): 1457-1473.

[16] Wu, S., Huang, Q., Zhao, L. Physics-guided deep learning-based inversion for airborne electromagnetic data. *Geophysical Journal International*, 2024, 238(3): 1774-1789.

[17] Cheng, K., Yang, X., Wu, X. Multiscale dilated convolutional neural networks for transient electromagnetic inversion. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 1-11.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)

© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261145

Lightweight Chinese Medicinal Material Recognition via Coordinate Attention and Knowledge Distillation

Shuhui MIN [a], Shaoyang ZHENG [a] Tianwei LEI [a,1] and Qinqin WU [a]

[a] *School of Big Data, Jiangxi Institute of Fashion Technology, Nanchang, Jiangxi, China*

Abstract. Automatic recognition of Chinese medicinal materials under limited computational budgets requires a balance between accuracy and complexity. We study a five-class pipeline combining MobileNetV3-small, Coordinate Attention (CA), and knowledge distillation (KD) from a frozen EfficientNet-B0 teacher; only the student is retained at inference. In the primary experiment, MobileNetV3-CA-KD obtains 85.83% test accuracy and 84.67% macro-F1 with 55.80 M FLOPs on a balanced 600-image subset. Its accuracy is 1.67 points below ShuffleNetV2 with approximately 61.2% fewer FLOPs, indicating an accuracy–computation trade-off rather than accuracy leadership. A paired single-seed ablation shows no additive CA benefit after distillation. Across three seeds, accuracy is $83.61\% \pm 1.92\%$, indicating non-negligible small-data variability.

Keywords. Chinese medicinal materials, lightweight network, MobileNetV3, coordinate attention, knowledge distillation

1. Introduction

Manual identification of Chinese medicinal materials is difficult to scale, while image-based recognition for sorting must often operate under limited computational budgets. Lightweight networks can struggle with dried materials that differ subtly in color, texture, and morphology. This motivates an empirical study of established CA [1] and KD [2] techniques within a MobileNetV3-small [3] pipeline.

The contributions are: (i) a fixed-seed, class-balanced five-class benchmark with duplicate controls; (ii) validation-guided CA placement; (iii) training-only KD from EfficientNet-B0 with validation-selected hyperparameters and no teacher cost at inference; and (iv) analysis of accuracy–computation trade-offs, CA placement, KD settings, training variability, and a paired ablation isolating KD from CA. This is an application-oriented evaluation of established components rather than a new attention or distillation mechanism.

[1] Corresponding Author: Tianwei Lei, e-mail: leitianwei@jift.edu.cn.

Table 1. Dataset statistics for the five medicinal material categories.

Class	Raw	Usable	Selected	Train	Val	Test
Codonopsis	207	177	120	72	24	24
Licorice	155	150	120	72	24	24
Goji berry	1259	1024	120	72	24	24
Astragalus	359	358	120	72	24	24
Honeysuckle	457	360	120	72	24	24
Total	2437	2069	600	360	120	120

2. Related Work

Lightweight networks and attention. MobileNetV3 [3] combines hardware-aware search with inverted residuals and squeeze-and-excitation (SE) [4]; ShuffleNetV2 [5] emphasizes practical efficiency; and EfficientNet-B0 [6] provides a mid-capacity reference. Coordinate Attention [1] retains directional positional cues at low cost, making CA placement relevant when compact networks distinguish visually similar categories.

Knowledge distillation. KD [2] transfers temperature-softened outputs to a compact student. Because the teacher is used only in training, it adds no parameters or FLOPs at inference; its benefit must nevertheless be established by controlled ablation.

Chinese medicinal material recognition. Chen et al. [7] studied 315 herbal categories, Miao et al. [8] used attention-augmented ConvNeXt, and Tian et al. [9] highlighted visually similar herbal-fruit categories. Wang et al. [10] combined channel and spatial attention for Chinese herbal slices; Gang et al. [11] proposed lightweight CCN-Net with MDCA drawing on Coordinate and Triplet Attention; Zheng et al. [12] applied enhanced KD to MobileNetV3-small and reported mobile deployment. Hou et al. [13] combined distillation and cross-attention with a ShuffleNet student. Our study uses validation-guided CA placement, a single logit-based teacher, and explicit accuracy–computation and small-data limitations.

3. Dataset

Collection and cleaning. The raw corpus contained 2,437 images from a pre-existing five-class collection: Codonopsis, Licorice, Goji berry, Astragalus, and Honeysuckle. Pillow verified readability; SHA-256 removed within-class exact duplicates and screened cross-class conflicts, while perceptual hashing identified near-duplicate groups. Removing 368 exact duplicates left 2,069 usable images; no corrupted images or cross-class exact-duplicate conflicts were found.

Sampling and split. With seed 42, a class-balanced, diversity-aware procedure selected 120 images per class while limiting repeated perceptual-hash groups. The 600 images were split by near-duplicate group at 6:2:2 (72/24/24 per class), so exact or near duplicates never crossed subsets. Only 600 of 2,069 usable images enter the experiment, leaving substantial data unexploited. Table 1 summarizes the statistics and Figure 1 shows examples.



Figure 1. Representative samples from the five medicinal material categories.

Metrics. We report test accuracy and macro-F1. For the primary experiments, CA placement, KD hyperparameters, and epoch selection are based on validation macro-F1; test results are used only for final evaluation.

4. Method

The proposed model in the primary experiment, denoted **MobileNetV3-CA-KD**, uses CA after `features[8]` and `features[12]` of torchvision MobileNetV3-small (CA[8,12]) and an EfficientNet-B0 teacher for KD. Only the student is retained at inference (Figure 2).

4.1. CA-Enhanced MobileNetV3-small Student

MobileNetV3-small [3] is adapted to five logits. CA [1] pools along two spatial directions to encode channel and positional information. Identity-preserving residual gates are inserted at the selected stages so the initialized feature path retains the pretrained representation while learning the attention contribution.

The insertion position of CA materially affects performance. We compare three configurations—CA[3,8,12], CA[8,12], and CA[12]—selecting by validation macro-F1 (Section 5.2). **CA[8,12]** achieves the best validation macro-F1 and is adopted.

4.2. Knowledge Distillation Framework

The student is trained under KD with EfficientNet-B0 [6] as the teacher, minimizing

$$\mathscr{L} = (1-\alpha)\,\mathscr{L}_{\mathrm{CE}}(y, \sigma(z_s)) + \alpha T^2\,\mathrm{KL}\big(\sigma(z_t/T)\,\big\|\,\sigma(z_s/T)\big)\,, \tag{1}$$

where z_t and z_s are teacher and student logits, T is temperature, and α weights the soft-target term. A validation-only grid over $T \in \{2,4,6\}$ and $\alpha \in \{0.3,0.5\}$ selects $T = 4$, $\alpha = 0.3$ (Section 5.3); the test set is not used for tuning. The frozen teacher remains in evaluation mode and is discarded after training, so KD adds no teacher parameters or FLOPs at student inference.

4.3. Training Details

All models use ImageNet initialization, 224×224 input, batch size 32, and AdamW with backbone/new-layer learning rates $10^{-4}/10^{-3}$ and weight decay 10^{-4}. General CE

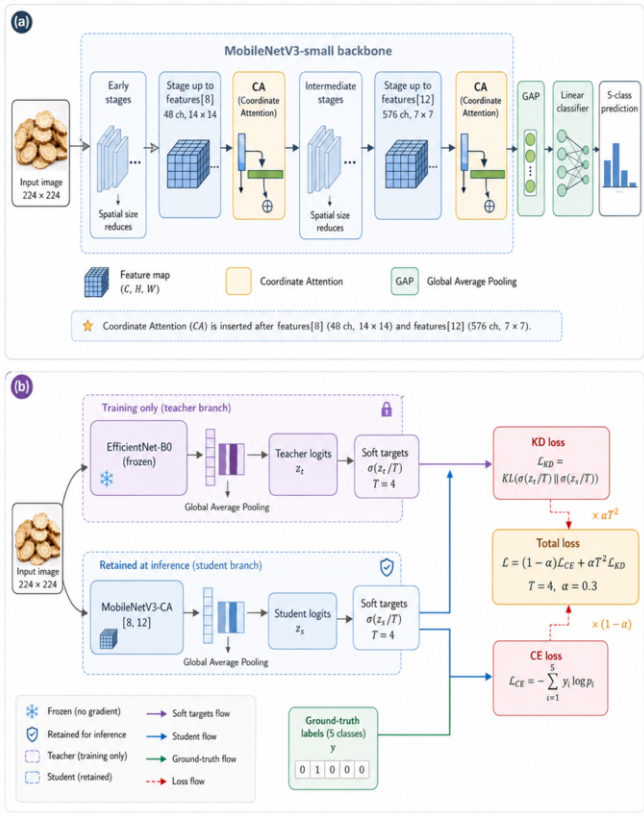


Figure 2. Overview of the proposed method.

training uses random resized crop $[0.8, 1.0]$, horizontal flip, $\pm 15^\circ$ rotation, color jitter, and ImageNet normalization; validation/test use resize 256 and center crop 224. The backbone is frozen for epochs 1–3 and unfrozen at epoch 4. Baselines use at most 50 epochs/patience 8; the CA ablation uses 60/patience 10.

KD training additionally uses label smoothing 0.1, crop $[0.9, 1.0]$, a cosine schedule with three warm-up epochs, and a student backbone frozen only in epoch 1. Runs use at most 50 epochs/patience 10, $T = 4$, $\alpha = 0.3$, and validation macro-F1 selection; the teacher stays frozen in evaluation mode. Unless stated otherwise, seed 42 is used.

5. Experiments

5.1. Main Results

All models use the same training, validation, and test split; model-specific training settings follow Section 4.3. Table 2 compares three off-the-shelf architectures, the CA-enhanced student, and the proposed model.

In the primary comparison, MobileNetV3-CA-KD records test accuracy and macro-F1 8.33 and 8.71 points above the separately trained MobileNetV3 baseline. Because

Table 2. Primary comparison (seed 42).

Model	Acc. (%)	Macro-F1 (%)	Params (M)	FLOPs (M)
MobileNetV3-small	77.50	75.96	1.523	55.49
ShuffleNetV2 ×1.0	87.50	86.72	1.259	143.89
EfficientNet-B0	85.00	83.57	4.014	384.54
MobileNetV3 + CA[8,12]	81.67	80.56	1.557	55.80
Proposed MobileNetV3-CA-KD	85.83	84.67	1.557	55.80

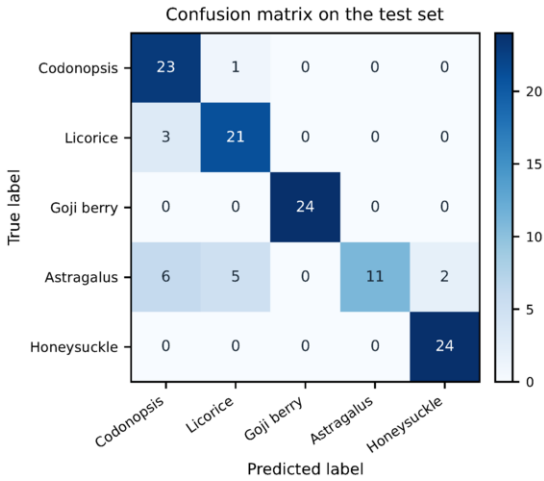


Figure 3. Confusion matrix of the proposed MobileNetV3-CA-KD on the test set.

CE and KD variants use different training recipes, this gap is not an isolated estimate of CA or KD effects. Its 85.83% accuracy is 1.67 points below ShuffleNetV2 with approximately 61.2% fewer FLOPs, so the result is an accuracy–computation trade-off rather than accuracy leadership. Against EfficientNet-B0, the 0.83-point difference on 120 test images is only a small point estimate; the student uses about 61.2% fewer parameters and 85.5% fewer FLOPs. Figure 3 shows class-wise predictions.

5.2. Ablation on CA Insertion Positions

The three CA placements yield (Val Macro-F1/Test Acc./Test Macro-F1) of 0.8305/0.8000/0.7848 for [3,8,12], 0.8474/0.8167/0.8056 for [8,12], and 0.8123/0.8250/0.8013 for [12]. Thus [8,12] is selected strictly by validation macro-F1; the one-sample higher test accuracy of [12] does not affect selection.

5.3. Additional Ablation and Robustness Analysis

Table 3 reports the validation-only KD grid from the original experiments. $T = 4$, $\alpha = 0.3$ attains the best validation macro-F1 (0.8472) and is adopted; no grid run evaluates the test set, so test data are not used for hyperparameter tuning.

Table 3. Validation-only KD grid from the original experiments: best validation macro-F1 (best epoch).

	$\alpha = 0.3$	$\alpha = 0.5$
$T = 2$	0.8389 (6)	0.8398 (6)
$T = 4$	**0.8472** (5)	0.8391 (5)
$T = 6$	0.8387 (5)	0.8219 (5)

Table 4. Supplementary reconstructed comparison (seed 42).

Configuration	Val Macro-F1	Test Acc. (%)	Test Macro-F1 (%)
Plain MobileNetV3 CE	0.8413	80.83	79.94
MobileNetV3 + CA CE	0.8329	81.67	80.82
MobileNetV3 + KD	0.8508	88.33	87.18
MobileNetV3 + CA + KD	0.8245	85.00	84.44

Historical checkpoints needed for the four-row supplementary comparison were not all retained, so Table 4 reports independent runs reconstructed from the recovered protocols. Its values should not replace the original runs in Table 2. The comparison was conducted after the primary configuration was fixed to examine CA additivity under KD and was not part of the original model-selection search space. The distilled variants used the same reconstructed teacher and identical KD settings, but each was initialized from its corresponding reconstructed CE checkpoint; these starting checkpoints differed in validation macro-F1 (0.8413 vs. 0.8329).

The reconstructed CE runs give mixed evidence: CA has lower validation macro-F1 (0.8329 vs. 0.8413) but slightly higher test metrics. Because these CE runs retain model-specific schedules and KD uses a different training recipe, cross-recipe CE-to-KD gaps are descriptive rather than isolated KD effects. In the directly matched distilled pair, KD-only has higher validation and test performance than CA+KD; thus, no additive CA benefit is observed in this single-seed reconstruction. This does not establish that CA is generally detrimental.

Across seeds 42, 123, and 2026, the proposed model obtains 85.83/84.67, 82.50/81.03, and 82.50/80.47% accuracy/macro-F1, respectively, yielding $83.61\% \pm 1.92\%$ accuracy and $82.06\% \pm 2.28\%$ macro-F1. These results indicate favorable average performance with non-negligible small-data variability.

6. Discussion and Limitations

Primary confusion-matrix errors concentrate on Astragalus: 11/24 samples are correct, while 6 are predicted as Codonopsis, 5 as Licorice, and 2 as Honeysuckle. Visual similarity in color, texture, and morphology may contribute; future work should target hard examples and stronger fine-grained representations. Generalization is limited by five categories, a 120-image test set, and use of only 600 of 2,069 usable images. Future evaluation should use more of the available data with class weighting or resampling and other appropriate imbalance-aware losses, repeated splits or stronger cross-validation, broader categories, and external validation. Parameters and FLOPs are complexity proxies; real-device latency, FPS, memory footprint, and power consumption remain unmeasured.

7. Conclusion

In the primary experiment, MobileNetV3-CA-KD attains 85.83% accuracy and 84.67% macro-F1 at 55.80 M FLOPs, giving an accuracy–computation trade-off with student-only inference. A supplementary single-seed reconstructed comparison shows no additive CA benefit after distillation, while cross-recipe CE-to-KD differences are not isolated KD effects. Conclusions remain limited by the balanced subset, small test set, seed variability, and lack of real-device measurements; future work should use more data, repeated splits, external validation, and constrained-device evaluation.

Acknowledgments

This work was supported by the Special Funded Bidding Project and the University-level Student Research Project of Jiangxi Institute of Fashion Technology (Grant Nos. JFM-BZX2543 and JF-LX-XSZX-202537).

References

[1] Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design. In: CVPR; 2021. p. 13713–13722.
[2] Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network. arXiv:1503.02531, 2015.
[3] Howard A, Sandler M, Chu G, Chen LC, Chen B, Tan M, et al. Searching for MobileNetV3. In: ICCV; 2019. p. 1314–1324.
[4] Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: CVPR; 2018. p. 7132–7141. doi: 10.1109/CVPR.2018.00745.
[5] Ma N, Zhang X, Zheng HT, Sun J. ShuffleNet V2: Practical guidelines for efficient CNN architecture design. In: ECCV; 2018. p. 116–131.
[6] Tan M, Le QV. EfficientNet: Rethinking model scaling for convolutional neural networks. In: ICML; 2019. p. 6105–6114.
[7] Chen W, Tong J, He R, et al. An easy method for identifying 315 categories of commonly-used Chinese herbal medicines based on automated image recognition using AutoML platforms. Inform Med Unlocked. 2021;25:100607.
[8] Miao J, Huang Y, Wang Z, Wu Z, Lv J. Image recognition of traditional Chinese medicine based on deep learning. Front Bioeng Biotechnol. 2023;11:1199803.
[9] Tian D, Zhou C, Wang Y, et al. NB-TCM-CHM: Image dataset of Chinese herbal medicine fruits and its application in classification through deep learning. Data Brief. 2024;54:110405.
[10] Wang J, Mo W, Wu Y, Xu X, Li Y, Ye J, Lai X. Combined channel attention and spatial attention module network for Chinese herbal slices automated recognition. Front Neurosci. 2022;16:920820.
[11] Gang H, Sheng G, Wang X, Jiang J. CCNNet: a novel lightweight convolutional neural network and its application in traditional Chinese medicine recognition. J Big Data. 2023;10:114.
[12] Zheng L, Long W, Yi J, Liu L, Xu K. Enhanced knowledge distillation for advanced recognition of Chinese herbal medicine. Sensors. 2024;24(5):1559.
[13] Hou Q, Yang W, Liu G. Chinese herbal medicine recognition network based on knowledge distillation and cross-attention. Sci Rep. 2025;15:1687.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261146

Research on an Intelligent Detection System for Municipal Manhole Covers Based on Edge-Cloud Collaboration

Haotian Luo, Zheng Sheng[1], Shengming Liang, and Mingdi Li
School of Information Engineering, Huzhou University, Huzhou 313000, China

Abstract. This study proposes an edge-cloud collaborative intelligent detection architecture for municipal manhole covers. A lightweight YOLO11n model is deployed on the edge side to achieve rapid preliminary screening of manhole cover anomalies. Suspected abnormal images are then optimized and uploaded to the cloud server, where a YOLO11 instance segmentation model performs refined damage analysis. Experiments are conducted on a self-constructed dataset. The edge-side YOLO11n model maintains a high recall rate while ensuring real-time performance, effectively preventing missed detections. Meanwhile, the cloud-side YOLO11 segmentation model achieves high detection accuracy in pixel-level identification of defects such as manhole cover displacement, loss, and breakage. Furthermore, the system integrates with a visual monitoring platform that supports real-time warning and decision analysis. This research not only addresses the challenge of balancing detection accuracy and efficiency in resource-constrained scenarios but also provides an effective solution for the intelligent operation and maintenance of municipal drainage systems, demonstrating the potential of integrating deep learning with edge computing in civil engineering applications.

Keywords: YOLO11, Manhole cover detection, Deep learning, Municipal Manhole Cover.

1. Introduction

As a critical component of urban infrastructure, the safe and stable operation of municipal drainage networks directly affects urban living conditions, flood control capability, and overall infrastructure sustainability [1]. Manhole covers, serving as key access nodes within drainage systems, are continuously exposed to complex environmental conditions and are therefore prone to defects such as cracks, displacement, and loss. Traditional manual inspection methods suffer from low efficiency, high labor costs, and significant safety risks. With the rapid development of smart cities and the widespread adoption of Internet of Things (IoT) technologies, intelligent detection has become an inevitable trend for improving pipeline network management [2]. However, most existing deep learning-based detection methods rely on centralized cloud processing, which often leads to poor real-time performance and excessive bandwidth consumption. The integration of edge computing with artificial intelligence provides a promising solution to these challenges. By distributing computational tasks to edge

[1] Corresponding Author, Sheng Zheng, School of Information Engineering, Huzhou University, Huzhou 313000, China; E-mail: 03083@zjhu.edu.cn.

devices, efficient and low-latency intelligent perception can be achieved [3]. In response to the practical demand for detecting defects in municipal manhole covers, this study proposes an edge-cloud collaborative intelligent detection architecture aimed at overcoming the limitations of traditional approaches and supporting intelligent operation and maintenance of urban drainage systems.

In recent years, computer vision technologies have achieved remarkable progress in civil infrastructure inspection, and deep learning based object detection models have been widely applied to defect identification in bridges, pavements, and road facilities [4]. Models such as YOLOv5 and YOLOv8 demonstrate excellent real-time detection performance. However, their computational complexity limits deployment in resource-constrained edge environments [5]. Meanwhile, image segmentation models such as U-Net and DeepLabV3+ enable pixel-level defect analysis but require substantial computational resources [6]. From a system architecture perspective, research on edge computing has gradually advanced. Although embedded platforms such as NVIDIA Jetson support lightweight model deployment, most existing studies focus on optimizing individual models rather than designing end-to-end collaborative systems [7]. Specifically for manhole cover detection, current methods exhibit three main shortcomings. First, model redundancy often leads to insufficient real-time performance on edge devices. Second, it remains difficult to balance detection accuracy with computational efficiency. Third, deep integration between IoT sensing, detection models, and visualization platforms is often lacking [8]. These limitations highlight the necessity of developing an edge-cloud collaborative detection system.

From a deployment-oriented perspective, a clear research gap exists in designing an end-to-end edge-cloud collaborative inspection system that simultaneously ensures high-recall screening at the edge, high-precision pixel-level refinement in the cloud, and scalable engineering implementation for smart city infrastructure management [9].

To address these challenges, this study proposes an intelligent manhole cover detection system based on YOLO11. A lightweight YOLO11n model is deployed on the edge side to perform high-recall preliminary screening of manhole cover anomalies. Subsequently, a YOLO11 instance segmentation model on the cloud side conducts fine-grained analysis of suspected anomalies, providing pixel-level defect information. An IoT-based architecture and visualization platform is integrated to support end-to-end monitoring and decision-making. The main contributions of this work are summarized as follows:

(1) An edge-cloud collaborative detection mechanism is designed to effectively balance detection latency and accuracy.

(2) An end-to-end system from data acquisition to early warning is implemented, facilitating practical deployment.

(3) A manhole cover dataset containing 4288 images across four condition categories is established, covering diverse real-world road scenarios and providing high-quality data support for model training and evaluation.

The remainder of this paper is organized as follows. Section 2 introduces the system architecture and methodology. Section 3 presents experimental results and performance evaluation. Section 4 discusses practical application value. Finally, conclusions and future work are provided in Section 5.

2. System Design and Implementation

2.1. Overall System Architecture

The proposed system adopts a hierarchical edge-cloud collaborative architecture consisting of a perception layer, an edge layer, a cloud layer, and an application layer, enabling end-to-end integration of data acquisition, processing, analysis, and visualization (see Figure 1) [10]. The perception layer is deployed at municipal manhole cover locations and includes IoT devices such as high-definition cameras for real-time image acquisition [11]. The edge layer is implemented on a Raspberry Pi 5 platform and runs a lightweight YOLO11n model to perform real-time analysis of video streams, preliminarily filtering suspected abnormal manhole covers. Through an optimized transmission strategy, only images containing suspected anomalies are uploaded to the cloud layer [12].

The cloud layer is supported by a high-performance GPU server, where a YOLO11 instance segmentation model performs pixel-level defect identification on uploaded images and generates detailed inspection results [13]. The application layer provides a web-based visualization platform that supports real-time monitoring, historical data querying, and early warning notifications. By distributing computational tasks across layers, the proposed architecture significantly reduces bandwidth consumption while maintaining a balance between detection accuracy and real-time performance.

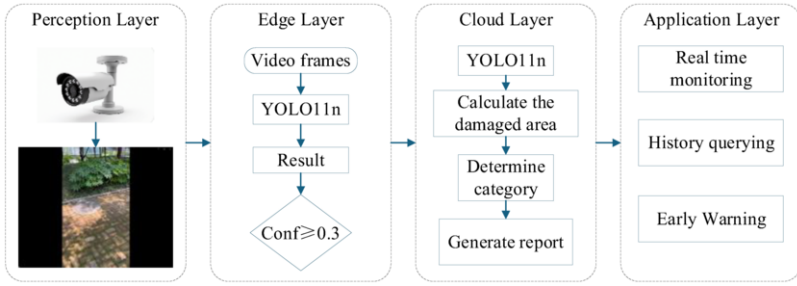


Figure 1. Workflow of Edge-Cloud Systems.

2.2. Core Algorithm Design

The algorithm design emphasizes collaboration between edge-side and cloud-side models (see Figure 2). On the edge side, the YOLO11n model is employed for preliminary anomaly screening. Its lightweight design is based on a CSPDarknet backbone and a PANet neck, incorporating depthwise separable convolutions to reduce parameter count. During training, data augmentation and anchor box re-clustering are applied to improve model adaptability [14]. The model outputs bounding boxes and confidence scores. A dynamic confidence threshold is adopted, and detections with confidence scores greater than 0.3 are regarded as suspected anomalies [11], ensuring a high-recall strategy and minimizing missed detections.

On the cloud side, a YOLO11 instance segmentation model is deployed. By integrating multi-scale feature fusion and attention mechanisms, the model enables fine-grained, pixel-level identification of manhole cover defects such as damage, displacement, and loss. The EIoU loss function is employed to improve bounding box regression accuracy by directly modeling absolute width and height errors, which is

particularly suitable for localizing elliptical manhole cover contours caused by viewpoint variations. The edge-cloud collaboration strategy incorporates anomaly filtering and transmission optimization to reduce overall resource consumption.

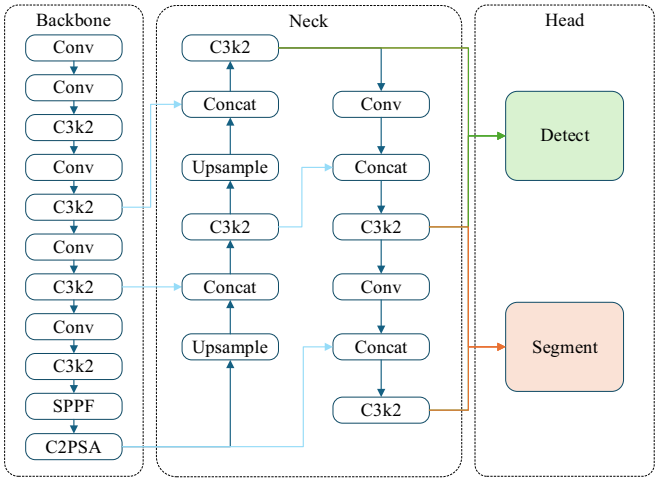


Figure 2. Architecture of the YOLO11 Object Detection and Instance Segmentation Models.

2.3. Engineering Implementation

The engineering implementation covers hardware configuration, software development, and deployment workflow. On the edge side, Raspberry Pi 5 embedded devices equipped with cameras are deployed at manhole cover sites. On the cloud side, a server equipped with an NVIDIA T4 GPU (16 GB memory) ensures high-performance computation.

Software development is based on the PyTorch framework. Edge-side applications are deployed using ONNX Runtime, supporting hybrid CPU/GPU inference for real-time video processing. Cloud-side services are implemented via RESTful APIs for receiving uploaded images and returning segmentation results. The visualization platform is developed using Vue 3, Element Plus, and ECharts, providing map-based displays, detection statistics, and alarm management. Models are trained offline and then encapsulated into edge and cloud components using containerization technologies to ensure scalability and maintainability.

3. Experimental Results and Analysis

3.1. Experimental Setup

A comprehensive experimental environment was constructed to validate the proposed edge-cloud collaborative detection system. The dataset was collected from real municipal manhole covers in the drainage network of Huzhou City, covering different time periods and weather conditions to ensure diversity. A total of 4288 high-resolution images were collected and annotated using Labelme.

The experimental architecture follows the proposed hierarchical design. Raspberry Pi 5 devices were deployed on the edge side, while a server equipped with an RTX 3060

GPU was used on the cloud side (see Figure 3). The software environment is based on PyTorch, with ONNX Runtime used for edge-side inference acceleration.



(a) Webcam (b) Raspberry Pi 5

Figure 3. Experimental Environment and Equipment.

Table 1 summarizes the key training parameters used for both the edge-side detection model and the cloud-side segmentation model. These parameters were determined based on convergence stability, hardware resource constraints, and deployment requirements. Specifically, the learning rate and optimizer were selected to ensure smooth gradient updates and robust generalization, while the batch size and input resolution were configured according to GPU memory limitations and real-time inference considerations. The dataset partitioning follows a standard training-validation-test split to ensure objective performance evaluation. Overall, the parameter configuration aims to balance detection accuracy, computational efficiency, and practical deployability within the proposed edge-cloud collaborative framework.

Table 1. Training Parameters.

Parameter	Value
batch	16
lr0	0.0001
optimizer	AdamW
imgsz	640×640
epochs	200
data partitioning	Training set: 3223, Validation set: 805, Test set: 260

Eight evaluation metrics are employed: mAP@0.5, mAP@0.5:0.95, Precision, Recall, F1-score, Parameters, Time, and GFLOPs [15].

The metric mAP is the average of the Average Precision (AP) across all categories, serving as the core evaluation metric in object detection. It comprehensively measures the model's localization accuracy and classification precision. The calculation equation is as follows:

$$mAP=\frac{\sum_{i=1}^{n} AP_i}{n} \tag{1}$$

Among these, n represents the total number of categories, and AP_i denotes the average precision of category i. The metric mAP@0.5 denotes the mean Average Precision calculated with an IoU threshold set to 0.5. An IoU threshold of 0.5 implies that a predicted bounding box is considered a true positive if it overlaps with the ground-truth bounding box by more than 50%. This metric imposes a relatively lenient requirement on localization accuracy and is generally used for evaluating a model's basic detection capability. In contrast, mAP@0.5:0.95 represents the average mAP computed over multiple IoU thresholds from 0.5 to 0.95 with a step size of 0.05. This metric enforces stricter localization criteria and offers a more comprehensive assessment of detection performance across varying precision levels. It is adopted as the official benchmark in the COCO dataset and serves as a core evaluation metric for YOLO11.

Precision is the proportion of true positive samples among all predicted positive samples. The equation is as follows:

$$\text{Precision}=\frac{T_P}{T_P+F_P} \tag{2}$$

Where, T_P represents true positives, and F_P represents false positives.

Recall is the proportion of true positive samples successfully predicted among all actual positive samples. The equation is as follows:

$$\text{Recall}=\frac{T_P}{T_P+F_N} \tag{3}$$

where F_N represents false negatives.

The F1-score is the harmonic mean of Precision and Recall, providing a balanced evaluation of the model's ability to minimize false positives and false negatives. The equation is as follows:

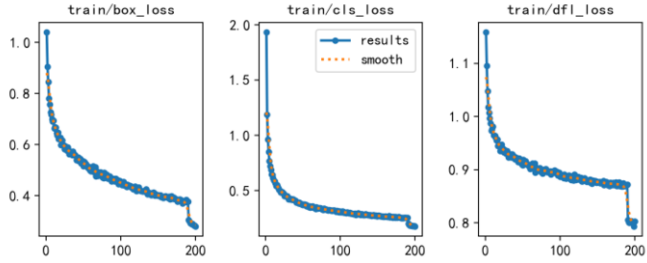
$$F1=2\times\frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \tag{4}$$

Parameters refer to the total number of trainable parameters in the model, reflecting its complexity and storage requirements. Time denotes the time required for the model to process a single image.

3.2. Results Analysis

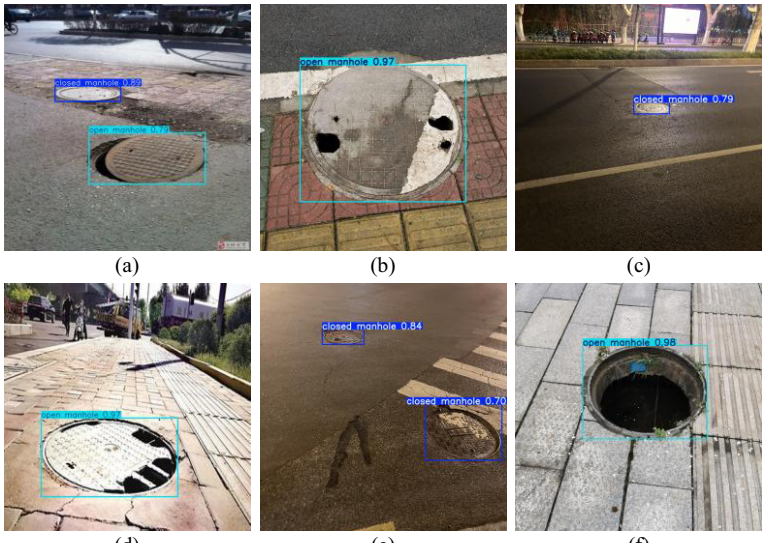
The lightweight YOLO11n model deployed at the edge exhibits excellent performance in the preliminary anomaly screening task. Evaluation results on the test set indicate that the model achieves a recall of 91.2%, ensuring a very high detection rate of abnormal manhole covers. Meanwhile, while maintaining high recall, the precision reaches 97.7%, implying that most of the uploaded suspected anomalies are true anomalies, which effectively reduces unnecessary computational burden on the cloud side. The model converges rapidly during training, and the loss function stabilizes after approximately 50 epochs, demonstrating that the model architecture is well designed and sufficiently trained (see Figure 4).

In practical deployment, the edge-side model can effectively identify various types of anomalies and generate high-confidence bounding boxes, providing a reliable basis for subsequent upload decisions (see Figure 5).



(a) box loss (b) cls loss (c) dfl loss

Figure 4. Training loss curves of the object detection model over 200 epochs.

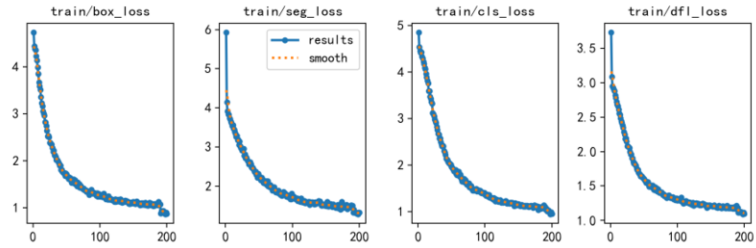


(a) (b) (c)

(d) (e) (f)

Figure 5. Examples of Edge-Side Detection Results.

After receiving the suspected anomaly images uploaded from the edge side, the YOLO11 segmentation model deployed in the cloud performs pixel-level fine-grained analysis. The training curves indicate that the YOLO11 segmentation model exhibits favorable convergence behavior during training, with validation metrics steadily improving as the number of epochs increases and reaching a plateau after approximately 100 epochs, demonstrating that the model has sufficiently learned the data characteristics (see Figure 6).

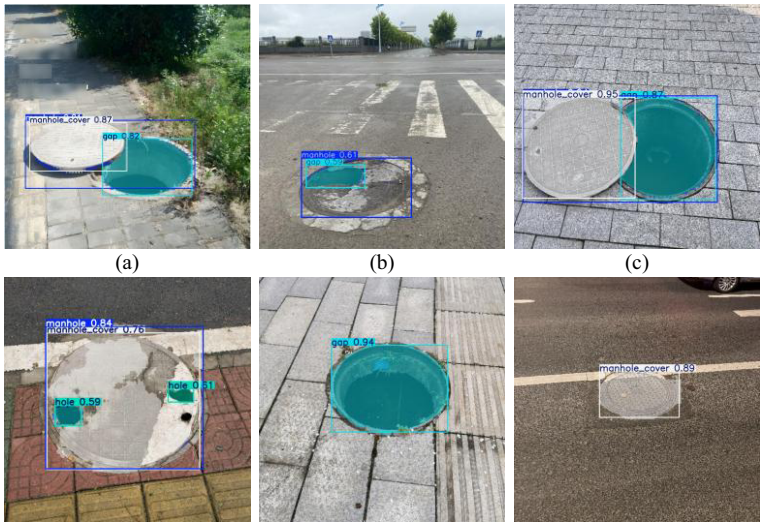


(a) box loss (b) seg loss (c) cls loss (d) dfl loss

Figure 6. Training loss curves of the instance segmentation model over 200 epochs.

The instance segmentation model demonstrates stable performance in delineating fine-grained features such as damage region contours and defect edges. It is capable of distinguishing local structures including manhole cover gaps and cracks, while maintaining accurate mask fitting under complex background conditions (see Figure 7).

The instance segmentation model shows clear advantages in pixel-level analysis tasks by providing detailed information on the area and contours of damaged regions. However, it incurs relatively high inference and computational costs, making it more suitable for deployment on the server side.



(a) (b) (c)

(d) (e) (f)

Figure 7. Examples of Cloud-Side Detection Results.

3.3. Discussion

The experimental results verify the practical feasibility and engineering value of the proposed edge-cloud collaborative detection framework. By combining high-recall edge screening with high-precision cloud segmentation, the system effectively balances computational efficiency and defect identification accuracy in municipal inspection scenarios.

From a system perspective, the edge-side YOLO11n model achieves a Recall of 91.2% and a Precision of 97.7%, ensuring that abnormal manhole covers are rarely missed

during preliminary screening while minimizing unnecessary cloud uploads. This high-recall strategy, implemented with a confidence threshold of 0.3, demonstrates that lightweight models can maintain strong detection capability under embedded deployment constraints.

On the cloud side, the YOLO11 instance segmentation model provides pixel-level defect delineation and achieves stable convergence after approximately 100 epochs. Compared with bounding-box-only detection, segmentation enables accurate structural damage interpretation, particularly for crack boundaries and displacement contours under complex backgrounds. Compared with traditional cloud-only processing architectures, the proposed collaborative mechanism reduces bandwidth transmission by filtering normal samples at the edge while preserving high defect detection reliability. Compared with edge-only deployment, the cloud segmentation stage significantly enhances fine-grained analysis capability. Therefore, the improvement does not lie solely in model accuracy but in system-level optimization through functional task decomposition.

However, certain limitations remain. The dataset is primarily collected from Huzhou City, which may influence cross-regional generalization performance. In addition, rare defect categories are relatively underrepresented, which may affect robustness in long-term deployment scenarios.

4. Conclusion

This study proposes and validates an intelligent municipal manhole cover inspection system based on an edge-cloud collaborative architecture. The framework successfully resolves the trade-off between real-time responsiveness and detection accuracy through hierarchical task decomposition. Quantitative results demonstrate that the edge-side YOLO11n model achieves a Recall of 91.2% and a Precision of 97.7%, converging within approximately 50 epochs while maintaining real-time inference performance on Raspberry Pi 5. The cloud-side YOLO11 instance segmentation model achieves stable convergence after approximately 100 epochs and provides accurate pixel-level delineation of defect regions, significantly enhancing structural damage interpretation capability.

The most robust finding of this study is that high-recall edge filtering combined with refined cloud segmentation enables reliable anomaly detection while substantially reducing bandwidth transmission and redundant computation. This confirms that collaborative deployment outperforms single-layer cloud-only or edge-only strategies in smart municipal inspection scenarios. Beyond algorithmic performance, the implemented end-to-end system integrating IoT perception, edge inference, cloud analysis, and visualization management demonstrates practical deployability and operational value. In pilot application, the system reduces hazard response time from traditional periodic inspection cycles to near real-time alerting, supporting proactive infrastructure maintenance.

Overall, the proposed framework provides a scalable and engineering-ready solution for intelligent urban drainage inspection and establishes a validated paradigm for edge-cloud collaborative vision systems in smart city infrastructure management.

Future research will focus on expanding multi-regional datasets to enhance cross-domain generalization, incorporating multimodal sensing data to improve contextual robustness, and exploring distributed collaboration among multiple edge nodes. In

addition, online continual learning mechanisms will be investigated to support long-term adaptive deployment in evolving urban environments.

References

[1] Zhou B, Zhao W, Guo W, Li L, Zhang D, Mao Q, Li Q. Smartphone-based road manhole cover detection and classification. Automation in Construction 140, (2022) https://doi.org/10.1016/j.autcon.2022.104344

[2] Mohamed A, Fouad MMM, Elhariri E, El-Bendary N, Zawbaa HM, Tahoun M, Hassanien AE: RoadMonitor: an intelligent road surface condition monitoring system. Intelligent Systems' 2014, pp. 377-387 (2015) https://doi.org/10.1007/978-3-319-11310-4_33

[3] Sheng Z, Chen F, Liu Q, Gao B, Zhang J, Zhao K, Liu Q, Zang Y: Real-time waterlogging monitoring on urban roads using edge computing. Water Resources Management 39, 5273-5287 (2025) https://doi.org/10.1007/s11269-025-04202-w

[4] Liu J, Zhao J, Cao Y, Wang Y, Dong C, Guo C: Road manhole cover defect detection via multi-scale edge enhancement and feature aggregation pyramid. Scientific Reports 15, (2025) https://doi.org/10.1038/s41598-025-95450-8

[5] Jiang P, Ergu D, Liu F, Cai Y, Ma BJPcs: A Review of Yolo algorithm developments. 199, 1066-1073 (2022) https://doi.org/10.1016/j.procs.2022.01.135

[6] Tian W, Chong K, Zhang J: DU-Net-L: an effective and lightweight segmentation model for alfalfa images that integrates the strengths of DeepLabV3+ and U-Net. aBIOTECH 6, 763-773 (2025) https://doi.org/10.1007/s42994-025-00235-2

[7] Sun Q, Li P, He C, Song Q, Chen J, Kong X, Luo Z: A lightweight and high-precision passion fruit YOLO detection model for deployment in embedded devices. Sensors 24, (2024) https://doi.org/10.3390/s24154942

[8] Mankotia A, Shukla AK: IOT based manhole detection and monitoring system using Arduino. Materials Today: Proceedings 57, 2195-2198 (2022) https://doi.org/10.1016/j.matpr.2021.12.264

[9] Hector R, Umar M, Mehmood A, Li Z, Bhattacharyya S: Scalable object detection for edge cloud environments. Frontiers in Sustainable Cities 3, (2021) https://doi.org/10.3389/frsc.2021.675889

[10] Sathupadi K, Achar S, Bhaskaran SV, Faruqui N, Abdullah-Al-Wadud M, Uddin J: Edge-cloud synergy for AI-enhanced sensor network data: a real-time predictive maintenance framework. Sensors (Basel) 24, (2024) https://doi.org/10.3390/s24247918

[11] Lei Z, Zhang Y, Wang J, Zhou M: Cloud-edge collaborative defect detection based on efficient Yolo networks and incremental learning. Sensors (Basel) 24, (2024) https://doi.org/10.3390/s24185921

[12] Karacaoglu B, Sahin ME: Optimized YOLO architectures for efficient Kiwi detection in precision agriculture on embedded systems. Sci. Rep. 15, 45113 (2025) https://doi.org/10.1038/s41598-025-32770-9

[13] Feiten RB, Freitas EPd, Oliveira MM: Comparative analysis of cloud-based AI object detection services and YOLO11: performance, cost, and usability evaluation. Journal of Cloud Computing (2026) https://doi.org/10.1186/s13677-025-00835-9

[14] Sapkota R, Karkee MJapa.: Ultralytics YOLO evolution: An overview of YOLO26, YOLO11, YOLOv8 and YOLOv5 object detectors for computer vision and pattern recognition. (2025) https://doi.org/10.48550/arXiv.2510.09653

[15] Wu M, Lin H, Shi X, Zhu S, Zheng B: MTS-YOLO: A multi-task lightweight and efficient model for tomato fruit bunch maturity and stem detection. Horticulturae 10, (2024) https://doi.org/10.3390/horticulturae10091006

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261147

Multi-Scale Dynamic Spatiotemporal Fusion for Traffic Flow Prediction

Zhiyong HE, and Xiongtao ZHANG[1]
School of Information Engineering, Huzhou University, Huzhou 313000, China

Abstract. Intelligent transportation systems rely on traffic flow forecasting to plan routes, manage transportation efficiently, and mitigate congestion. However, reliable prediction remains difficult in heterogeneous road structures because traffic data vary across multiple temporal scales and reflect dynamically changing spatial interactions among sensors. To address these issues, this paper proposes a Multi-Scale Dynamic Spatiotemporal Fusion model (MSDSTF). Within the framework, multi-scale temporal encoding is integrated with a dual-mask spatial self-attention module for modeling local and long-range spatial relations and a delay-aware transformation module for describing lag effects in spatial propagation. Across four real-world datasets, the proposed model achieves competitive forecasting accuracy, while ablation experiments verify the contribution of each major component.

Keywords: Traffic flow prediction, spatiotemporal modeling, multi-scale learning, dynamic graph neural networks, intelligent transportation systems

1. Introduction

Urbanization has increased the scale and complexity of transportation systems, thereby intensifying congestion and reducing operational efficiency. Within Intelligent Transportation Systems (ITS) [1], traffic flow prediction infers future traffic states from historical observations. The resulting forecasts support emergency response, congestion control, route guidance, and scheduling. Improving their accuracy and robustness is therefore important for reducing urban traffic pressure [2]. Despite these benefits, traffic forecasting still faces the fundamental challenge of accurately representing complex spatiotemporal dependencies as they evolve [3]. Deep learning approaches have consequently been explored for this task. Early work modeled spatial correlations with Convolutional Neural Networks (CNNs) and temporal dynamics with Recurrent Neural Networks (RNNs). Graph Neural Networks (GNNs) [4] subsequently gained attention because prediction models can use them to incorporate road network structure naturally [5].

Despite the strong performance of GNN-based approaches, several limitations persist [6]. First, travel demand and unexpected events cause spatial correlations in traffic systems to change over time, yet many methods continue to model them statically. This limitation has motivated multi-scale graph decoupling [7]. Second, geographically distant nodes can display similar traffic behavior when they serve comparable urban functions. Existing methods often represent these long-range semantic relations

[1] Corresponding Author, Xiongtao Zhang, School of Information Engineering, Huzhou University, Huzhou 313000, China; E-mail: 1047897965@qq.com.

inadequately [8]. Figure 1 illustrates the representative traffic-flow patterns considered here: dynamic variation, long-range correlation, propagation delay, and periodicity.

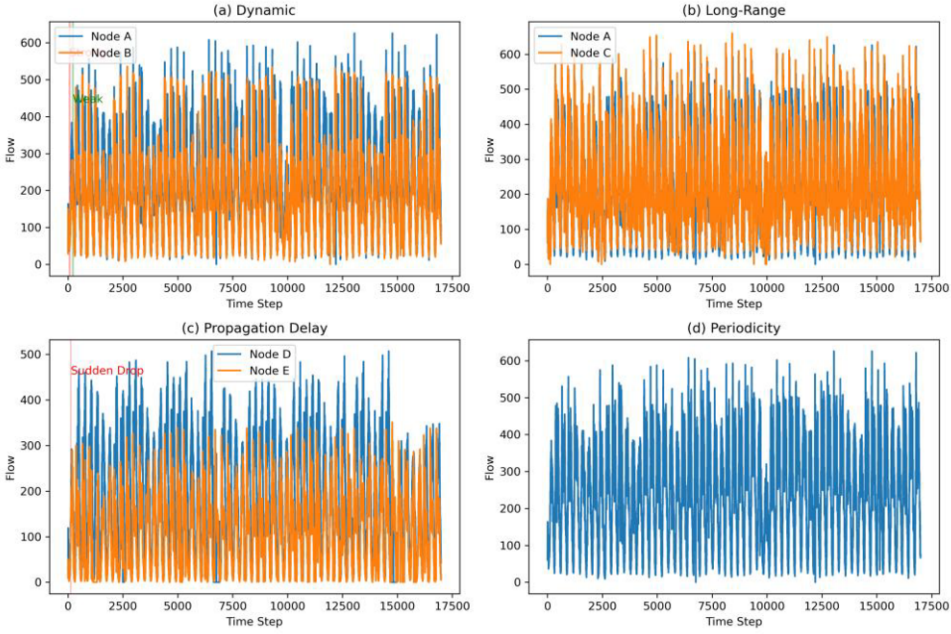


Figure 1. Findings in Traffic Flow Prediction.

Third, many graph-based models assume immediate message transmission, although traffic influence commonly propagates after a temporal lag. Self-supervised spatiotemporal learning [9] and progressive graph-Transformer modeling [10] have also been investigated in recent studies.

To address these issues, a Transformer-based framework is developed around multi-scale dynamic spatiotemporal fusion. Dynamic graph masks guide spatial self-attention within a unified process that represents local geographic and global semantic neighbors. A delay-aware feature transformation module then captures temporal lag during the spatial propagation of traffic information. Temporal self-attention further learns dynamic patterns in traffic data, providing a more complete description of their temporal evolution. The principal contributions are as follows:

(1) The proposed multi-scale dynamic spatiotemporal self-attention model improves traffic flow prediction by jointly representing delay-aware, dynamic, and long-range dependencies.

(2) Geographic proximity and semantic similarity are combined through a dual-mask spatial self-attention mechanism to model global and local spatial relations within a unified process.

(3) Temporal self-attention is integrated with a delay-aware feature transformation module to explicitly represent temporal lags during spatial information propagation.

On four real-world traffic datasets, the method delivers competitive forecasting performance; ablation analysis validates the contribution of the designed components.

2. Methods

Figure 2 presents the overall architecture of the proposed MSDSTF framework. Its three principal parts are the data embedding layer, stacked spatiotemporal fusion encoding layers, and the output layer. Each component is described below.

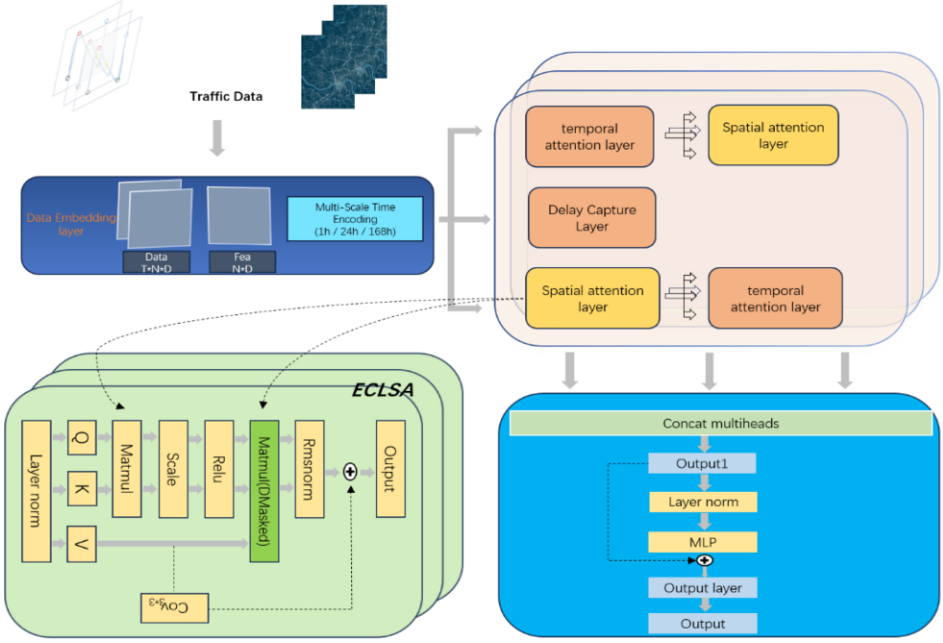


Figure 2. Overall Architecture of the MSDSTF Model.

2.1. Data Embedding Layer

The data embedding layer first uses a fully connected layer to project the raw input into a high-dimensional latent space. Here, T, N, and D represent the numbers of time steps, nodes, and embedding dimensions, respectively. A spatiotemporal embedding scheme is then constructed to retain network structure and temporal characteristics.

$$X_{data} \in R^{T \times D \times N} \tag{1}$$

Road network structure is represented in the spatial dimension through graph Laplacian embedding. Given adjacency matrix A, degree matrix D, and identity matrix I, the normalized graph Laplacian takes the form $\Delta = I - D^{-1/2} A D^{-1/2}$. The corresponding eigen-decomposition is written as $\Delta = U^T \Lambda U$, where Λ denotes the eigenvalue matrix and U the eigenvector matrix. After the k smallest non-trivial eigenvectors are selected and linearly projected, the spatial embedding $X_{spe} \in R^{N \times d}$ is obtained and retains the graph's global structural information. Traffic data exhibits multi-scale periodic patterns along the temporal dimension. Specifically, each timestamp t is mapped to a weekly index $t_w(t) \in \{1,\ldots,7\}$ and a daily index $t_d(t) \in \{1,\ldots,1440\}$, which are embedded into vectors X_w

and X_d ($X_w, X_d \in R^{T\times d}$) respectively. Relative temporal positions are represented by a temporal positional encoding $X_{tpe} \in R^{T\times d}$. Following the MMM method, fine-grained temporal dependencies are learned through accurate modeling of temporal positions. Let $t = \{1,2,\ldots,T\}$; the d-dimensional positional encoding assigned to each time step is defined as:

$$X_{tpe}(i) = \begin{cases} sin\left(t/10000^{2i/d}\right), if\ is\ even \\ cos\left(t/10000^{2(i-1)/d}\right), if\ is\ odd \end{cases} \tag{2}$$

All components are summed element by element to obtain the final embedding:

$$X_{emb} = X_{data} + X_{spe} + X_w + X_d + X_{tpe} \tag{3}$$

The resulting representation X_{emb} subsequently enters the spatiotemporal encoder and, for simplicity, is denoted by X throughout the remaining sections.

2.2. Enhanced Spatial Self-Attention

The spatiotemporal encoding layer uses self-attention to represent complex dependencies that change over time and contains three components. Its spatial self-attention component includes geographic and semantic sub-modules for local and long-range spatial relationships. The geographic branch also contains a delay-aware feature transformation module that explicitly represents temporal lags in spatial information propagation. Finally, temporal self-attention learns dynamic patterns over long temporal ranges.

Given an input tensor X, spatial self-attention represents time-dependent interactions among traffic nodes. For every time step, node features are transformed into query, key, and value matrices, from which pairwise spatial attention scores are calculated. Normalized attention scores then weight the value matrix to produce the spatial self-attention output.

$$X \in R^{T\times D\times N} X_{t::} \in R^{T\times D} X_{::n} \in R^{T\times D} \tag{4}$$

$$Q_t^{(S)} = X_{t::}W_Q^S, K_t^{(S)} = X_{t::}W_K^S, V_t^{(S)} = X_{t::}W_V^S \tag{5}$$

where W_Q^S , W_K^S , W_V^S denote learnable parameters, whereas d' specifies the dimensionality of the query, key, and value matrices. Spatial attention scores are calculated as:

$$A_t^{(S)} = \frac{Q_t^{(S)}\left(K_t^{(S)}\right)^T}{\sqrt{d'}} \tag{6}$$

where $A_t^{(S)} \in R^{N\times N}$ represents spatial dependencies among nodes. Because traffic states change over time, these dependencies also vary temporally. The spatial self-attention output is therefore expressed as:

$$ESSA\left(Q_t^{(S)}, K_t^{(S)}, , V_t^{(S)}\right) = softmax\left(A_t^{(S)}\right)V_t^{(S)} \tag{7}$$

Under this formulation, the spatial graph is fully connected and every pair of nodes interacts. Many such interactions, however, provide little information for traffic prediction. A geographic mask and a semantic mask are therefore applied to suppress noisy dependencies while jointly representing short- and long-range spatial relations. Physical distance determines the geographic mask, which retains nearby node pairs. Sequence similarity defines the semantic mask and links distant nodes that exhibit similar traffic patterns. Historical sequences are compared with DTW before training to construct the semantic relations, thereby avoiding repeated pairwise DTW calculations during training. When the two masks are applied to the attention matrix, uninformative interactions are suppressed and informative spatial relations receive greater emphasis.

$$M_{geo}M_{sem}M_{geo}M_{sem} \tag{8}$$

$$GESSA\left(Q_t^{(S)}, K_t^{(S)}, , V_t^{(S)}\right) = softmax(\left(A_t^{(S)} \odot M_{geo}\right)V_t^{(S)} \tag{9}$$

$$SESSA\left(Q_t^{(S)}, K_t^{(S)}, V_t^{(S)}\right) = softmax\left(A_t^{(S)} \odot M_{sem}\right)V_t^{(S)} \tag{10}$$

2.3. Enhanced Delay-aware Feature Transformation

The effects of congestion and local incidents spread progressively through real traffic systems rather than occurring instantaneously. The delay-aware feature transformation module describes this propagation delay by extracting latent information from short-term historical traffic sequences and injecting it into the geographic spatial attention branch. A sliding window first generates short-term sequences, which the kShape algorithm clusters into representative temporal patterns. The current sequence of each node is then projected into a feature space and compared against the resulting prototypes. These similarities form a delay-aware representation that is added to the spatial key matrix to model the delay explicitly.

$$p_i P = \{p_i\}_{i=1}^{N_p} x_{(t-S+1:t),n} \tag{11}$$

$$u_{(t,n)} = x_{(t-S+1:t),n}W_u \tag{12}$$

An embedding is obtained for every pattern centroid as follows:

$$m_i = p_i W_m \tag{13}$$

The similarity between $u_{(t,n)}$ and m_i is computed as:

$$w_i = softmax\left(u_{(t,n)}m_i\right) \tag{14}$$

The delay-aware representation is subsequently calculated by:

$$r_{(t,n)} = \sum_{i=1}^{N_p} w_i \left(p_i W_c \right) \tag{15}$$

The spatial Key matrix is finally updated as:

$$K_t^{(S)} = K_t^{(S)} + R_t \tag{16}$$

2.4. Enhanced Temporal Self-Attention

Traffic states observed at different time steps are strongly correlated along the temporal dimension, reflecting periodic trends and long-term variations. The form of these temporal relations can vary across nodes and traffic conditions. An enhanced temporal self-attention module therefore learns the resulting dynamic and heterogeneous temporal patterns adaptively from traffic data. $X(:,n,:)$

$$\left(Q_n^{(T)}, K_n^{(T)}, V_n^{(T)} \right) = X(:,n,:) \left(W_Q^{(T)}, W_K^{(T)}, W_V^{(T)} \right) \tag{17}$$

where $W_Q^{(T)}, W_K^{(T)}, W_V^{(T)} \in R^{d \times d'}$ are learnable parameters. The temporal attention scores are subsequently calculated as:

$$A_n^{(T)} = \frac{Q_n^{(T)} \left(K_n^{(T)} \right)^T}{\sqrt{d'}} \tag{18}$$

Different nodes can therefore attend to different historical steps according to their individual traffic patterns. In addition, the global receptive field of temporal self-attention supports the representation of long-range temporal dependency.

$$ESTA\left(Q_n^{(T)}, K_n^{(T)}, V_n^{(T)} \right) = softmax\left(A_n^{(T)} \right) V_n^{(T)} \tag{19}$$

3. Experimental Results and Analysis

3.1. Experimental Settings and Results

Four public traffic datasets are used to evaluate MSDSTF: the graph-based highway datasets PeMS04, PeMS07, and PeMS08, together with the grid-based NYTaxi dataset. The three graph-based datasets are divided into training, validation, and test sets at a ratio of 6:2:2. Their task uses the previous 12 time steps to predict the next 12 time steps. For NYTaxi, a 7:1:2 split is adopted, and the previous 6 time steps provide the input for predicting inflow and outflow at the next step. Before training, all input data are normalized by Z-score. Performance is measured using MAE, MAPE, and RMSE; for NYTaxi, these metrics are averaged over inflow and outflow. Unless stated otherwise, results are averaged over repeated runs. The baseline set comprises GMAN [11],

ASTGNN [12], PDFormer [13], Cy2Mixer [14]. Tables 1, 2, 3, and 4 report the corresponding comparisons for PeMS04, PeMS07, PeMS08, and NYTaxi, respectively.

Table 1. Model Performance on PeMS04.

Models	GMAN	ASTGNN	PDFormer	Cy2Mixer	ours
MAE	19.139	18.601	18.321	18.144	18.133
MAPE	13.192	12.630	12.103	11.933	12.093
RMSE	31.601	31.028	29.965	30.022	30.018

Table 2. Model Performance on PeMS07.

Models	GMAN	ASTGNN	PDFormer	Cy2Mixer	ours
MAE	20.967	20.616	19.832	19.453	19.433
MAPE	9.052	8.861	8.529	8.114	8.303
RMSE	34.097	34.017	32.870	32.892	32.641

Table 3. Model Performance on PeMS08.

Models	GMAN	ASTGNN	PDFormer	Cy2Mixer	ours
MAE	15.307	14.974	13.583	13.531	13.515
MAPE	10.134	9.489	9.046	8.862	9.067
RMSE	24.915	24.710	23.505	23.223	23.235

Table 4. Model Performance on NYTaxi.

Models	GMAN	ASTGNN	PDFormer	Cy2Mixer	ours
MAE	13.270	12.978	12.364	3.801	3.796
MAPE	13.893	13.647	12.781	29.023	15.693
RMSE	21.661	21.189	20.167	5.375	5.367

3.2. Ablation Study

To assess how individual MSDSTF components contribute to performance, the full model is compared with four simplified variants. (1) In w/ GCN, a Graph Convolutional Network (GCN) replaces spatial self-attention (SSA), preventing the model from capturing dynamic or long-range spatial dependency. (2) The w/o Mask variant removes both mask matrices and allows every node to attend to all other nodes. (3) The w/o Multi-Scale variant omits multi-scale temporal encoding and no longer represents traffic patterns at different temporal scales. (4) The w/o Delay variant removes the delay-aware feature transformation module and disregards delayed spatial propagation. $M_{geo}M_{sem}$

Figure 3 summarizes the following ablation findings:

1. In modeling dynamic and long-range spatial dependencies, SSA performs better than GCN;
2. Performance decreases when the mask design is removed, indicating that prediction benefits from identifying important node interactions and representing both global and local spatial relations.
3. Compared with the full model, w/o Multi-Scale performs worse, supporting the value of multi-scale temporal modeling.

4. Performance also declines for w/o Delay, showing that delayed spatial propagation contributes to traffic forecasting.

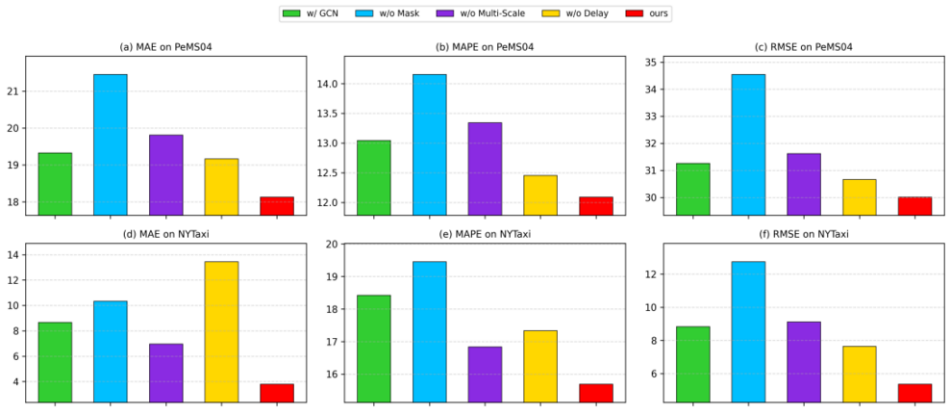


Figure 3. Ablation Study on PeMS04 and NYTaxi.

4. Conclusion

This study presents MSDSTF, a traffic flow prediction model based on spatiotemporal fusion self-attention. The model combines dual-mask spatial self-attention and delay-aware feature transformation with multi-scale spatiotemporal encoding to capture richer dependency patterns in traffic data. Competitive forecasting performance is obtained on four public datasets, and ablation results validate the main components. Future work can incorporate richer external factors and construct semantic relations more efficiently for traffic forecasting.

References

[1] Yin X, Wu G, Wei J, Shen Y, Qi H, and Yin B: Deep learning on traffic prediction: methods, analysis, and future directions. IEEE Transactions on Intelligent Transportation Systems 23(6), 4927–4943 (2022). https://doi.org/10.1109/TITS.2021.3054840

[2] Tedjopurnomo DA, Bao Z, Zheng B, Choudhury FM, and Qin AK: A survey on modern deep neural network for traffic prediction: trends, methods and challenges. IEEE Transactions on Knowledge and Data Engineering 34(4), 1544–1561 (2022). https://doi.org/10.1109/TKDE.2020.3001195

[3] Wang J, Jiang J, Jiang W, Li C, and Zhao WX: LibCity: An open library for traffic prediction. In: Proceedings of the 29th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, pp. 145–148. ACM (2021). https://doi.org/10.1145/3474717.3483923

[4] Zhang J, Zheng Y, and Qi D: Deep spatio-temporal residual networks for citywide crowd flows prediction. In: Proceedings of the AAAI Conference on Artificial Intelligence 31(1), 1655–1661 (2017). https://doi.org/10.1609/aaai.v31i1.10735

[5] Li Y, Yu R, Shahabi C, and Liu Y: Diffusion convolutional recurrent neural network: data-driven traffic forecasting. In: International Conference on Learning Representations (2018). https://openreview.net/forum?id=SJiHXGWAZ

[6] Yu B, Yin H, and Zhu Z: Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting. In: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, pp. 3634–3640 (2018). https://doi.org/10.24963/ijcai.2018/505

[7] Li H, Liu W, and Chen H: Multi-scale graph-decoupling spatial–temporal network for traffic flow forecasting in complex urban environments. Electronics 15(3), 495 (2026). https://doi.org/10.3390/electronics15030495

[8] Choi J, Choi H, Hwang J, and Park N: Graph neural controlled differential equations for traffic forecasting. In: Proceedings of the AAAI Conference on Artificial Intelligence 36(6), 6367–6374 (2022). https://doi.org/10.1609/aaai.v36i6.20587

[9] Li Z, Hu Z, Han P, Gu Y, and Cai S: SSL-STMFormer: self-supervised learning spatio-temporal entanglement Transformer for traffic flow prediction. In: Proceedings of the AAAI Conference on Artificial Intelligence 39(11), 12130–12138 (2025). https://doi.org/10.1609/aaai.v39i11.33321

[10] Chen L: PGSFormer: traffic flow prediction based on joint optimization of progressive graph convolutional networks with subseries transformer. Scientific Reports 16, 7200 (2026). https://doi.org/10.1038/s41598-026-35643-x

[11] Zheng C, Fan X, Wang C, and Qi J: GMAN: a graph multi-attention network for traffic prediction. In: Proceedings of the AAAI Conference on Artificial Intelligence 34(1), 1234–1241 (2020). https://doi.org/10.1609/aaai.v34i01.5477

[12] Guo S, Lin Y, Wan H, Li X, and Cong G: Learning dynamics and heterogeneity of spatial-temporal graph data for traffic forecasting. IEEE Transactions on Knowledge and Data Engineering 34(11), 5415–5428 (2022). https://doi.org/10.1109/TKDE.2021.3056502

[13] Jiang J, Han C, Zhao WX, and Wang J: PDFormer: propagation delay-aware dynamic long-range Transformer for traffic flow prediction. In: Proceedings of the AAAI Conference on Artificial Intelligence 37(4), 4365–4373 (2023). https://doi.org/10.1609/aaai.v37i4.25556

[14] Lee M, Choi YY, Park SW, Lee S, Ko J, and Hong J: Enhancing topological dependencies in spatio-temporal graphs with cycle message passing blocks. In: Proceedings of the Third Learning on Graphs Conference, PMLR 269, 33:1–33:17 (2025). https://proceedings.mlr.press/v269/lee25a.html

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)

© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261148

Narcissistic Rivalry Effect on Employee Voice Behavior under Abusive Supervision: A Moderated Mediation Model

Yunxia SU[a,b,1], Feng LI [a,b]
[a] *Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China;*
[b] *Department of Psychology, University of Chinese Academy of Sciences, Beijing 100049, China*

Abstract. Built on the dual-pathway model of narcissism, which distinguishes admiration and rivalry, and trait activation theory, this study tested how narcissistic rivalry drives employee voice behavior. Analyses drew on 247 valid questionnaires: confirmatory factor analysis (SPSS 26.0, AMOS 28.0), then hierarchical regression and the PROCESS 3.5 macro for the hypotheses. Narcissistic rivalry predicted voice negatively, relational avoidance motivation mediated the link, and abusive supervision moderated the first-stage path in the expected direction. The moderated mediation itself was not significant: the indirect effect stayed comparable across levels of abusive supervision.

Keywords. Narcissistic Admiration and Rivalry Concept (NARC), Trait Activation Theory (TAT), Relational Avoidance Motivation, Abusive Supervision, Narcissistic Rivalry, Voice Behavior

1. Introduction

Narcissus, the mythical figure absorbed by his own reflection, has become a lasting image of narcissism. Centuries later, Salvador Dalí returned to the same image in "Metamorphosis of Narcissus", while Sigmund Freud sought to understand narcissism as a psychological phenomenon involving the self and the unconscious. Their approaches differed, but they shared a basic question: what lies behind excessive investment in the self? The present study approaches this old question with modern personality and organizational research tools.

Contemporary research has shown that narcissism is not a single, uniform trait. In particular, Back et al. [1] distinguished narcissistic admiration from narcissistic rivalry. Whereas admiration is associated with the pursuit of attention and social recognition, rivalry reflects a more defensive and competitive response when one's status or self-image is threatened [1,2]. In simple terms, admiration seeks the spotlight, whereas rivalry is concerned with who else is standing in it. A discussion panel may even become a mutual admiration club until someone else receives more applause. Voice behavior—

[1] Corresponding author: 34895324@qq.com

extra-role communication in which employees offer constructive suggestions [3]—aims at improvement rather than criticism [4].

Despite this progress, several limitations remain. The cross-sectional design limits causal inferences, and reliance on self-report questionnaires may introduce bias. Future research could incorporate longitudinal data and multi-source assessments to validate these findings. Additionally, cultural factors specific to the Chinese context may influence generalizability, warranting cross-cultural studies. This study addresses these gaps by distinguishing narcissistic rivalry from admiration, testing relational avoidance motivation as a mediating mechanism, and examining abusive supervision as a situational boundary condition.

Problem formulation. Existing research has examined narcissism as an individual difference and voice behavior in relation to leaders and workplace conditions. Less attention has been given to whether employees' own narcissistic rivalry affects their willingness to speak up. The problem addressed in this study is whether narcissistic rivalry is associated with reduced voice behavior and, if so, whether relational avoidance motivation provides a psychological explanation for this association. A further question is whether abusive supervision changes the strength of the rivalry–avoidance relationship.

Main contribution of the present study. This study integrates employee personality and work environment factors, offering insights valuable for organizational psychology and management. It examines narcissistic rivalry as an antecedent of voice behavior, tests relational avoidance motivation as the mediating mechanism, and considers abusive supervision as a situational boundary condition. These findings inform organizational strategies to foster open communication and manage narcissistic traits effectively, extending the dual-pathway view of narcissism developed by Back et al. [1] and applying the trait-activation perspective to voice behavior. The main empirical contribution is to identify relational avoidance as a pathway through which narcissistic rivalry may inhibit employees' willingness to speak up.

Along this path, we introduce abusive supervision as a situational moderator—defined by Tepper [5] as subordinates' perception of sustained hostile verbal and nonverbal behavior displayed by their managers. As a typical negative-feedback situation, it may activate the defensive patterns of narcissistic rivalry and enlarge their effect on behavior. The overall model therefore combines "narcissistic rivalry–relational avoidance motivation–voice behavior" with the "abusive supervision" context: we ask how rivalry reaches voice behavior through the mediating mechanism, and where abusive supervision draws the boundary.

2. Research Hypotheses

2.1 Narcissistic Rivalry and Relational Avoidance Motivation

The dual-pathway model assigns the same goal—a positive self—to two forms of narcissism that operate in opposite directions [1,6]. Admiration reaches outward for social approval. Rivalry holds its ground by avoiding social failure: reputation is guarded vigilantly, impressions are managed, and "self-protection" strategies keep risk and self-exposure to a minimum [7,8]. Notably, Alicke and Sedikides [7] described self-protection as a response to threats to a positive self-view, while Baumeister et al. [8] linked self-presentational motives to individual differences in self-esteem.

In motivation theory, "avoiding harm" is the root of avoidance motivation; once activated, it tilts decision-making toward the safety and benefits of self-protection [9]. Carried into interpersonal life, this logic defines relational avoidance motivation as a self-protective response: people withdraw from closeness to keep their self-image and public face intact [7].

When rivals sense their self-image is at risk, negative schemas may become more influential: power struggles, devaluation of others, emotional detachment—a defensive, antagonistic pattern [1,10]. Add self-centeredness, distrust, the need to control, and competitiveness, and the result is systematic avoidance: intimacy is evaded, dependence resisted, and deeper emotional exchange avoided [11].

H1: Narcissistic rivalry should be positively related to relational avoidance motivation.

By proposing that narcissistic rivalry is positively related to relational avoidance motivation, this hypothesis aims to deepen understanding of personality influences on workplace behavior and encourage scholarly engagement.

2.2 The Mediating Effect of Relational Avoidance Motivation

LePine and Van Dyne [3] treated voice as discretionary communication through which employees can suggest improvements or raise work-related concerns. Van Dyne and LePine [4] further distinguished voice from other extra-role behaviors and examined its construct and predictive validity. Voice therefore gives employees a way to challenge the status quo, point out problems, and propose development [3]. Its benefits are well documented, yet for rival individuals a failed attempt to speak up may be experienced as a threat signal.

H2: Relational avoidance motivation should be negatively related to voice behavior.

Taken together, H1 and H2 suggest that rivalry may increase relational avoidance motivation, which in turn may reduce voice. Relational avoidance motivation thus appears to carry rivalry's effect onto voice behavior, in line with the idea of a transmission effect [9].

H3: Narcissistic rivalry should influence voice behavior through relational avoidance motivation.

2.3 The Situational Moderating Effect of Abusive Supervision

The dual-pathway framework also stresses that traits do not translate into behavior in a one-way manner; situational factors interact with traits at every step [2]. Tepper [12] described abusive supervision as a form of sustained hostile behavior by supervisors, making it a plausible situational cue in which defensive tendencies may become more salient. If so, situational cues may activate or amplify the rivalry–avoidance chain described above.

Under low abusive supervision, rival individuals already treat interaction as costly. Once the image of self appears demeanable, self-protection intensifies to fend off harm and hold onto a positive self-view [7,8]. Krizan and Johar [13] showed how such protection can slide into rivalry—rejection, strained relations, low popularity, criticism, and lost trust follow. Context, though, is not fixed. In the stressor–strain model [14], strain may become more widespread as abuse increases; individual differences then

explain less of the behavior, because under strong abuse almost everyone is strained and the "distinctiveness" of rivalry fades. Attribution theory [15] points in the same direction: strong abuse gives employees a clear external target, problems are blamed on the leader, and the rivalry–avoidance link loosens. We therefore expect:

H4: High abusive supervision should weaken the positive relationship between narcissistic rivalry and relational avoidance motivation.

Edwards and Lambert's [16] moderated path analysis framework draws the two hypotheses together. If a situational variable moderates the first-stage path linking an independent variable to its mediator, it should also moderate the strength of the indirect effect. Accordingly, we infer:

H5: The indirect effect of narcissistic rivalry on voice behavior via relational avoidance motivation will be weaker when abusive supervision is higher.

Figure 1 depicts the hypothesized model.

Within this framework, narcissistic rivalry is proposed to influence voice behavior indirectly through relational avoidance motivation, and this indirect pathway is expected to be moderated by the situational context of abusive supervision.

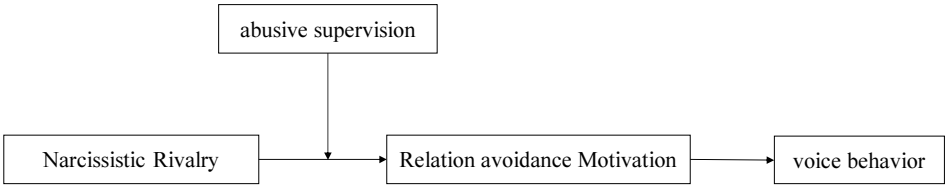


Figure 1. The moderated mediation model.

3. Participants and Methods

3.1 Participants

We first discarded questionnaires with missing responses, straight-line answering, or implausibly short completion times; we then screened the rest for missing values, outliers, and normality. No focal variable had missing values, and no case crossed the univariate outlier cutoff ($|z| > 3.29$) or the multivariate one (Mahalanobis distance, $p < 0.001$). For every study variable, skewness and kurtosis fell within acceptable bounds ($|\text{skewness}| < 2$, $|\text{kurtosis}| < 7$), suggesting that univariate normality was not seriously violated; these cutoffs are based on Kline [17]. The normality assumption is of limited concern here: with $N = 247$, the central limit theorem justifies the regression-based significance tests, and indirect effects were estimated via 5,000-sample bootstrapping in PROCESS, which does not require multivariate normality. After screening, 247 usable questionnaires remained—163 from women (66%) and 84 from men (34%). Participants were 34.16 years old on average (SD = 8.10), with an average organizational tenure of 7.87 years (SD = 6.67). The majority held a bachelor's degree (88%), and 12% held a master's degree or above.

3.2 Measures

Narcissistic rivalry was measured using the nine-item scale of Back et al. [1]; "I think most other people are failures" is a representative item. Employees self-reported, and Cronbach's α was acceptable.

Relational avoidance motivation was tapped by six items adapted from Ferris et al.'s [18] approach–avoidance measure, e.g., "I sometimes feel anxious in interpersonal interactions," capturing the urge to avoid or escape interpersonal contact; self-reports showed acceptable Cronbach's α. Given the adaptation, we ran a CFA in AMOS 28.0 on the six items. A one-factor model fit well, supporting unidimensionality; internal consistency was good (α = 0.825), and in the six-factor measurement model the items loaded on the relational avoidance factor with acceptable fit.

Abusive supervision was measured with Tepper's [5] fifteen-item scale (sample item: "My supervisor often ridicules me"), rated on a 7-point Likert scale (1 = strongly disagree, 7 = strongly agree); Cronbach's α was 0.863.

Voice behavior was assessed with the shortened scale of Liang et al. [19], which contains three promotive and three prohibitive items (e.g., "put forward constructive suggestions to improve the company's operation"; "dissuade, promptly, behaviors that undermine work efficiency"), rated on a 7-point Likert scale (1 = very untrue of me, 7 = very true of me). Cronbach's α was 0.804 for promotive voice, 0.837 for prohibitive voice, and 0.937 for the full scale.

4. Results

4.1 Descriptive Statistics and Correlation Analysis

Table 1 reports the descriptive statistics and correlation matrix for all focal variables, along with the demographic covariates (gender, age, tenure). Narcissistic rivalry (NR) correlated significantly and positively with relational avoidance motivation (AV) (r = 0.52, $p < 0.001$), suggesting these tendencies often co-occur. Relational avoidance motivation showed a significant negative association with voice behavior (r = -0.46, $p <$ 0.001), implying that voice behavior drops markedly among relationally avoidant employees. Narcissistic rivalry was likewise negatively correlated with voice behavior (r = -0.32, $p < 0.001$), and we tested the hypotheses.

Table 1. Means, standard deviations, and correlations among variables (N = 247)

Variable	M	SD	1	2	3	5	7	8	9
1. Gender	0.66								
2. Age	34.16	8.10	-0.13*						
3. Tenure	7.87	6.67	-0.14*	0.82***					
5. Narcissistic rivalry (T1)	3.16	1.12	-0.17**	0.17**	0.20**				
7. Relational avoidance motivation (T2)	3.44	1.11	-0.11	0.13*	0.13*	0.52***			
8. Voice behavior (T3)	5.33	0.98	-0.09	0.13*	0.07	-0.32***	-0.46***		
9. Abusive supervision (T1)	1.78	0.73	-0.10	0.01	-0.01	0.38***	0.27***	-0.27***	

Note: Gender coding: 0 = male, 1 = female; Cronbach's α coefficients are shown in parentheses on the diagonal; *** p < 0.001, ** p < 0.01, * p < 0.05 (two-tailed).

4.2 Common Method Bias Test

As all constructs were self-reported by employees, we examined common method bias with Harman's single-factor test. The first factor from an exploratory factor analysis of all items explained 32.30% of the total variance, well below the 40% threshold [20]; common method bias therefore does not appear to be a major concern here.

4.3 Hypothesis Testing

Hypotheses were evaluated using hierarchical regression in SPSS and the PROCESS 3.5 macro; psychological variables were standardized beforehand to avoid multicollinearity.

Narcissistic rivalry significantly and positively predicted relational avoidance motivation (Model 2: $\beta = 0.48$, $p < 0.001$, 95% CI = [0.35, 0.59]), supporting H1. When voice behavior served as the outcome and gender, age, and tenure were held constant, narcissistic rivalry again showed a significant negative relation with voice (Model 5: $\beta = -0.37$, $p < 0.001$, 95% CI = [-0.48, -0.24]). With relational avoidance motivation entered, the narcissistic rivalry–voice link stayed significant while shrinking in magnitude (Model 6: $\beta = -0.15$, $p < 0.05$, 95% CI = [-0.28, -0.03]); relational avoidance motivation, in turn, negatively predicted voice (Model 6: $\beta = -0.42$, $p < 0.001$, 95% CI = [-0.54, -0.29]), supporting H2.

We then ran Model 4 of PROCESS. The direct effect of narcissistic rivalry on voice remained significant (effect = -0.17, SE = 0.05, t = -3.80, $p < 0.001$). The indirect effect via relational avoidance motivation was -0.20 (SE = 0.05), with a 95% bootstrap confidence interval (5,000 samples) of [-0.30, -0.10] that excluded zero. H3 was supported: relational avoidance motivation partially mediated the narcissistic rivalry–voice relationship.

Table 2 reports the hierarchical regression results for the moderation model, showing how the narcissistic rivalry × abusive supervision interaction predicted relational avoidance motivation beyond the main effects.

Table 2. Hierarchical regression results for the moderating effect (N = 247)

Variable	Relational avoidance motivation (T2)			Voice behavior (T3)		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Gender	-0.09	-0.01	-0.02	-0.08	-0.13*	-0.14*
Age	0.06	0.06	0.04	0.22	0.23*	0.25**
Tenure	0.07	-0.01	0.00	-0.12	-0.06	-0.07
Narcissistic rivalry (T1)		0.48***	0.48***		-0.37***	-0.15*
Relational avoidance motivation (T2)						-0.42***
Abusive supervision (T1)		0.08	0.08			
Narcissistic rivalry (T1) × Abusive supervision (T1)			-0.11*			
R^2	0.03	0.28	0.29	0.03	0.16	0.28

ΔR^2	0.03	0.25	0.01	0.03	0.13	0.13
ΔF	2.31	42.62***	4.30*	2.30	36.48***	42.11***

Note: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. NR = narcissistic rivalry, AS = abusive supervision.

We then turned to the moderating role of abusive supervision. Before adding the interaction term, narcissistic rivalry positively predicted relational avoidance motivation (Model 2: $\beta = 0.48$, $p < 0.001$, 95% CI = [0.35, 0.59]). Once the interaction was included, it negatively predicted relational avoidance motivation (Model 3: $\beta = -0.11$, $p < 0.05$, 95% CI = [-0.22, -0.01]), supporting H4. Simple slopes computed at one standard deviation below and above the mean of abusive supervision revealed that narcissistic rivalry positively predicted relational avoidance motivation under both low ($\beta = 0.51$, $t = 7.49$, $p < 0.001$, 95% CI = [0.39, 0.66]) and high ($\beta = 0.41$, $t = 3.71$, $p < 0.001$, 95% CI = [0.17, 0.57]) abusive supervision; the positive effect was weaker, however, when abusive supervision was high.

Figure 2 visually illustrates the relationship between narcissistic rivalry and relational avoidance motivation under low and high abusive supervision, highlighting how the situational context differentially shapes the strength of this association.

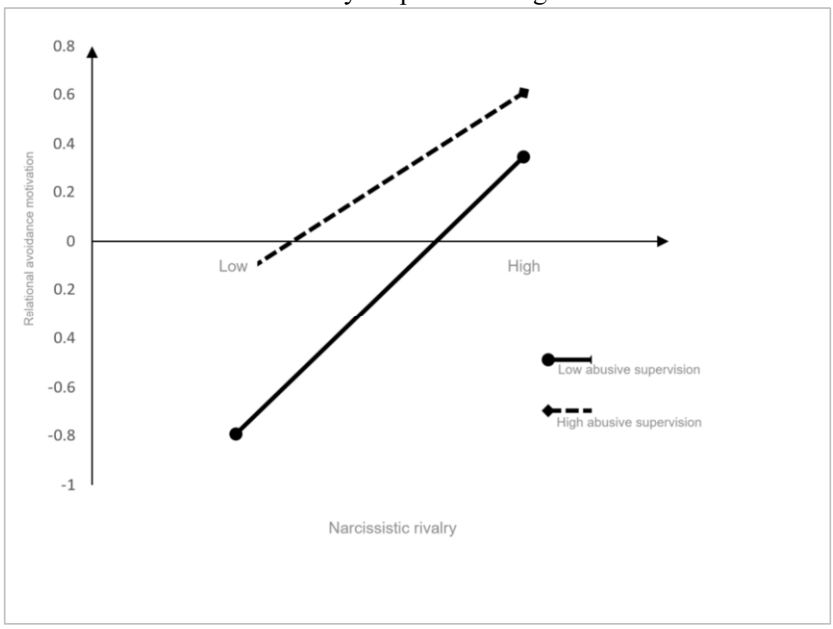


Figure 2. The moderating effect of abusive supervision on the relationship between narcissistic rivalry and relational avoidance motivation.

The moderated mediation was examined with Model 7 of PROCESS 3.5. Its index equaled 0.046 and was not significant; the 95% bootstrap CI (5,000 samples), [-0.02, 0.09], included zero. H5 was not supported. Whether the voice-suppressing effect of relational avoidance motivation is uniform across situations remains unclear; a restricted moderator range may also explain the null, since abusive supervision averaged only 1.78 (SD = 0.73) on a 7-point scale in this sample. Section 5 revisits this issue among other limitations.

5. Conclusions, Implications, and Limitations

This study addressed a straightforward question with a moderated mediation design: do employees' own traits shape their behavior at work? We proposed that narcissistic rivalry—a core personality trait [21]—activates relational avoidance motivation [22] and thereby suppresses voice behavior [23], with abusive supervision moderating the first-stage path and, through it, the overall indirect effect [16]. The data aligned with most of the story: rivalry predicted avoidance motivation positively; avoidance motivation predicted voice negatively and mediated the link; and the first-stage moderation behaved as expected, with the rivalry–avoidance link stronger under low abuse. Only the moderated mediation itself was not significant.

Two main contributions of this study are as follows:

1).It extended the dual-pathway model of narcissism. While Back et al. [1] separated admiration from rivalry, our data show rivalry reaching voice through relational avoidance, adding empirical support for the model's discriminant validity.

2).The proposed model integrated "employee traits" and "leaders' negative behavior" into one framework. While study [24] implied that situations activate trait expression, we tested the trait × situation interaction directly, answering calls to examine situational boundary conditions. This shifts attention within voice-inhibition research: instead of examining silence from the leader side [25], we started with employees' own traits and located relational avoidance motivation as one more psychological pathway that carries the suppression.

This study's findings have two practical implications:

1).Organizations should screen for abusive supervision [12] so that its amplifying effect is cut off early.

2).Employees high in rivalry need psychological safety interventions [26] that loosen relational avoidance motivation; only then can constructive voice and citizenship behavior surface in more domains.

However, the following six main limitations should be considered when interpreting these findings.

1).The three time points do not make the analysis longitudinal; causal direction remains to be pinned down by designs such as cross-lagged panel models.

2).Although the three-wave layout (T1: rivalry and abuse; T2: approach and avoidance motivation; T3: voice) should reduce common method bias, all constructs were self-reported, so such bias cannot be excluded [20]; anonymity and item ordering were used as procedural remedies, and multi-source ratings (supervisors or coworkers judging voice) would help further.

3).The data came from one industry; external validity is untested.

4).Abusive supervision was low on average ($M = 1.78$, $SD = 0.73$ on a 7-point scale; most participants had little such experience), which constrains the power and range of the moderation tests; generalizing the first-stage moderation and the moderated mediation to strongly abusive settings calls for caution.

5).The sample was skewed female (66%), and gender correlated with rivalry ($r = -0.17$); gender was controlled but not explored as a confound or moderator—a balanced sample would tell whether the model holds for both sexes.

6).Relational avoidance motivation rested on an adapted scale; CFA supported its factorial validity here, but validation across samples and occupations is still needed.

These limitations will be mitigated in our follow-up study, in particular, by adding the narcissistic admiration factor and widening the moderators.

References

[1] Back MD, Küfner AC, Dufner M, Gerlach TM, Rauthmann JF, Denissen JJ. Narcissistic admiration and rivalry: Disentangling the bright and dark sides of narcissism. J Pers Soc Psychol. 2013; 105(6): 1013-1037.
[2] Back MD, Küfner ACP, Leckelt M. Early impressions of grandiose narcissists: A dual-pathway perspective. In: Hermann AD, Brunell AB, Foster JD, editors. The handbook of trait narcissism: Key advances, research methods, and controversies. New York: Springer; 2018. p. 57-67.
[3] LePine JA, Van Dyne L. Predicting voice behavior in work groups. J Appl Psychol. 1998; 83(6): 853-868.
[4] Van Dyne L, LePine JA. Helping and voice extra-role behaviors: Evidence of construct and predictive validity. Acad Manage J. 1998; 41(1): 108-119.
[5] Tepper BJ. Consequences of abusive supervision. Acad Manage J. 2000; 43(2): 178-190.
[6] Wallace HM, Baumeister RF. The performance of narcissists rises and falls with perceived opportunity for glory. J Pers Soc Psychol. 2002; 82(5): 819-834.
[7] Alicke MD, Sedikides C. Self-enhancement and self-protection: What they are and what they do. Eur Rev Soc Psychol. 2009; 20(1):1-48.
[8] Baumeister RF, Tice DM, Hutton DG. Self-presentational motivations and personality differences in self-esteem. J Pers. 1989; 57(3): 547-579.
[9] Elliot AJ. The hierarchical model of approach-avoidance motivation. Motiv Emot. 2006; 30(2): 111-116.
[10] Grove JL, Smith TW, Girard JM, Wright AGC. Narcissistic admiration and rivalry: An interpersonal approach to construct validation. J Pers Disord. 2019; 33(6): 751-775.
[11] Kato T. The psychology of narcissism. Nanning: Guangxi Science & Technology Press; 2016. (in Chinese)
[12] Tepper BJ. Abusive supervision in work organizations: Review, synthesis, and research agenda. J Manage. 2007; 33(3): 261-289.
[13] Krizan Z, Johar O. Narcissistic rage revisited. J Pers Soc Psychol. 2015; 108(5): 784-801.
[14] Jex SM, Beehr TA. Emerging theoretical and methodological issues in the study of work-related stress. Res Pers Hum Resour Manag. 1991; 9: 311-365.
[15] Kelley HH. The processes of causal attribution. Am Psychol. 1973; 28(2): 107-128.
[16] Edwards JR, Lambert LS. Methods for integrating moderation and mediation: A general analytical framework using moderated path analysis. Psychol Methods. 2007; 12(1): 1-22.
[17] Kline RB. Principles and practice of structural equation modeling. 4th ed. New York: Guilford Press; 2016.
[18] Ferris DL, Johnson RE, Rosen CC, Djurdjevic E, Chang CH, Tan JA. When is success not satisfying? Integrating regulatory focus and approach/avoidance motivation theories to explain the relation between core self-evaluation and job satisfaction. J Appl Psychol. 2013; 98(2): 342-353.
[19] Liang J, Farh CIC, Farh JL. Psychological antecedents of promotive and prohibitive voice: A two-wave examination. Acad Manage J. 2012; 55(1): 71-92.
[20] Podsakoff PM, MacKenzie SB, Lee JY, Podsakoff NP. Common method biases in behavioral research: A critical review of the literature and recommended remedies. J Appl Psychol. 2003; 88(5): 879-903.
[21] Miller JD, Lynam DR, McCain JL, et al. Thinking structurally about narcissism: An examination of the Narcissistic Admiration and Rivalry Concept and its measurement. Assessment. 2016; 23(1): 3-9.
[22] Gable SL. Approach and avoidance social motives and goals. J Pers. 2006; 74(1):175-222.
[23] Morrison EW. Employee voice behavior: Integration and directions for future research. Acad Manage Ann. 2011; 5(1): 373-412.
[24] Tett RP, Burnett MF. A personality trait-based interactionist model of job performance. J Appl Psychol. 2003; 88(3): 500-517.
[25] Morrison EW, Milliken FJ. Organizational silence: A barrier to change and development in a pluralistic world. Acad Manage Rev. 2000; 25(4): 706-725.
[26] Edmondson A. Psychological safety and learning behavior in work teams. Adm Sci Q. 1999; 44(2): 350-383.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261149

ICT-Enabled Agricultural Systems for Food Security, Climate-Smart Agriculture, and Inclusive Rural Development

Rotimi-Williams BELLO[a, b, 1], Roseline Oluwaseun OGUNDOKUN[a], Pius A. OWOLAWI[a], Etienne A. van WYK[a], Chunling TU[a], and Odunayo IMORU[c]
[a] *Department of Computer Systems Engineering, Faculty of Information and Communication Technology, Tshwane University of Technology, South Africa*
[b] *Department of Mathematics and Computer Science, University of Africa, Toru-Orua, Nigeria*
[c] *School for Continuing Education, Red River College Polytechnic, Winnipeg, Manitoba, Canada*

Abstract. Traditional agricultural practices have been transformed by the rapid advancement of Information and Communication Technologies (ICT). The transformation, which involves data-driven, technology-enabled systems, is commonly referred to as e-Agriculture. This e-Agriculture incorporates state-of-the-art digital tools, including the Internet of Things (IoT), mobile technologies, Geographic Information Systems (GIS), cloud computing, big data analytics, and Artificial Intelligence (AI), to enhance agricultural resilience, productivity, and sustainability. This paper presents a comprehensive overview of e-Agriculture, including its main ICT components, key application areas, benefits, challenges, and future research directions. In this paper, special attention is given to how ICT-enabled agricultural systems support food security, climate-smart agriculture, and inclusive rural development, particularly in developing countries.

Keywords: Digital Transformation, e-Agriculture, ICT, Precision Farming, Smart Agriculture

1. Introduction

Agriculture is considered one of the vital elements of food security, job creation, and economic growth worldwide, especially in developing countries [1]. However, many serious challenges are affecting the agricultural sector, including limited resources, climate change, pests, population growth, and inefficiencies in the marketplace [2]. Addressing these challenges with traditional farming practices alone is a herculean task, necessitating the emergence and adoption of Information and Communication Technologies (ICT), through which agriculture has evolved into what is today referred to as e-Agriculture (a digitally connected ecosystem) [3-5]. e-Agriculture is defined by the Food and Agriculture Organization (FAO) as the employment of ICT for the

[1] Corresponding author, Rotimi-Williams BELLO, Department of Computer Systems Engineering, Faculty of Information and Communication Technology, Tshwane University of Technology, South Africa; E-mail: sirbrw@yahoo.com.

enhancement of farming practices and agricultural development [6]. The gap between farmers, researchers, policymakers, and markets is bridged by e-Agriculture through automation, timely and informed policies, and intelligent decision-making [7].

This paper explores the concept of e-Agriculture, examining its major ICT-driven applications and their roles in improving efficiency, productivity, and sustainability. The paper advances the existing body of knowledge with its novel conceptual integration, methodological contribution, regional focus and contextualization, and identification of research gaps and future directions.

2. Concept and Architecture of e-Agriculture

2.1. Definition of e-Agriculture

E-Agriculture is defined as the systematic use of ICT across the various value chains of agriculture, including production, processing, storage, distribution, and marketing [6, 8]. e-Agriculture encompasses technologies that support informed decision-making, including intelligent systems, sensor networks, mobile services, and digital platforms [9]. Figure 1 shows the concept and architecture of e-Agriculture.



Figure 1. Concept and architecture of e-Agriculture.

2.2. ICT Architecture for e-Agriculture

Four layers made up the architecture of a typical e-Agriculture, including (a) sensing layer, (b) communication layer, (c) data and processing layer, and (d) application layer. Figure 2 shows a typical e-Agriculture architecture, consisting of a sensing layer, a communication layer, a data and processing layer, and an application layer [10]. Table 1 presents real-world implementations and performance outcomes across the four-layer framework.

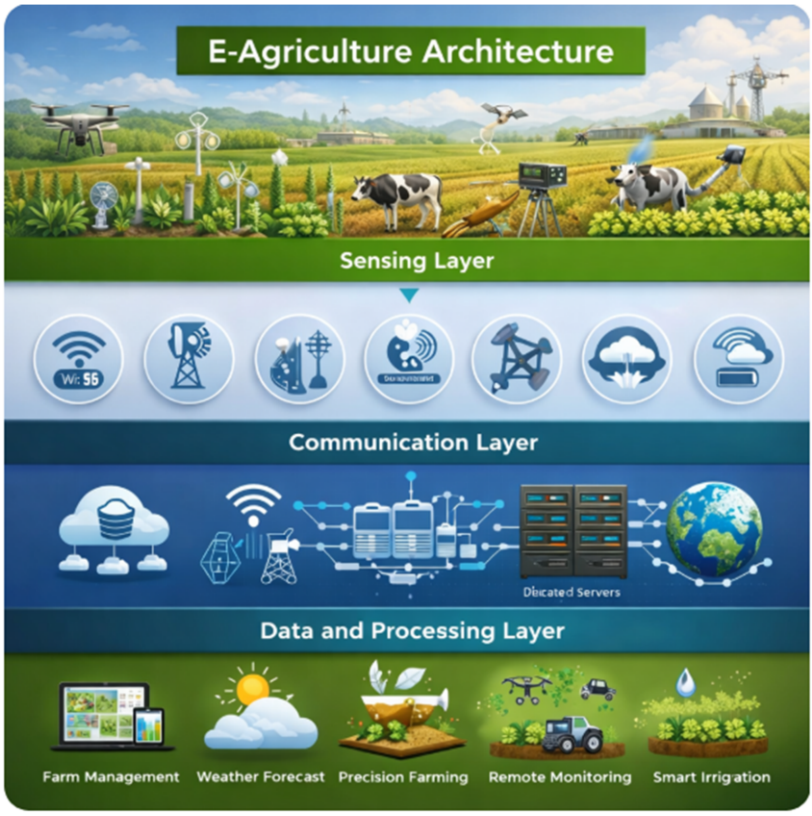


Figure 2. A typical e-Agriculture architecture, consisting of a sensing layer, a communication layer, a data and processing layer, and an application layer.

(a) Sensing Layer: This layer comprises sensors, IoT devices, satellite systems, and drones for data collection, such as temperature, soil moisture, and crop health.

(b) Communication Layer: This layer comprises wireless sensor networks, mobile networks (2G–5G), and the internet for the transmission of data.

(c) Data layer: This is also known as the "data and processing layer", and is made up of databases, big data analytics tools, and cloud services.

(d) Application layer: This consists of mobile applications, dashboards, advisory services, and decision support systems for the user.

Table 1. Real-World Implementations and Outcomes of the Four-Layer e-Agriculture Framework (with Emphasis on Developing Countries).

Framework Layer	Technologies Used	Real-World Implementation (Developing Countries)	Performance Metrics/ Outcomes	Key References
Sensing Layer	Soil moisture sensors, temperature/humidity sensors, remote sensing, IoT devices	IoT-based smart irrigation systems deployed in smallholder farms in Kenya and India	Up to 25–30% reduction in water usage; improved soil monitoring and irrigation timing	[11]
Communication Layer	GSM/3G/4G, LPWAN (LoRa, Sigfox), NB-IoT	Rural connectivity solutions using hybrid LPWAN–cellular networks in Sub-Saharan Africa	~20–30% reduction in communication costs; improved data transmission reliability in low-infrastructure settings	[12]

Data & Processing Layer	Machine Learning, Deep Learning, Cloud Computing, Edge Computing	AI-based crop yield prediction and disease detection systems applied in Nigeria and Bangladesh	15–25% improvement in yield prediction accuracy; reduced crop loss through early disease detection	[13, 14]
Application Layer	Mobile apps, decision support systems, farm management platforms	Mobile-based advisory platforms providing weather forecasts, irrigation scheduling, and market information (e.g., digital advisory services in East Africa)	Up to 20–25% increase in crop yield; improved farmer decision-making and income levels	[15]

3. Methodology

3.1. Review Approach

This study employs a narrative review method of conducting a structured and systematic literature search. This provides a foundation for both a conceptual synthesis of findings and interdisciplinary integration of results, while maintaining sound methodology. The primary aim of this review is to critically evaluate the current literature on the relationship between Artificial Intelligence (AI), Information and Communication Technologies (ICT), Renewable Energy Technologies, and smart agricultural systems.

3.2. Literature Search Strategy

Key scientific databases (Scopus, Web of Science, IEEE Xplore, Google Scholar) were used to conduct a comprehensive literature search on digital agriculture and AI-based systems, covering publications from 2010 to 2025. A comprehensive methodology was developed to incorporate various keyword combinations in conjunction with Boolean operators for each of the following phrases: "smart/precision agriculture", "artificial intelligence/machine learning", "ICT in agriculture", "renewable energy/energy systems", and "climate smart agriculture". Search terms were refined iteratively until desired keywords could be identified with sufficient specificity to yield an adequate set of records for future reference.

3.3. Study Selection Criteria

To ensure relevance and quality, the following inclusion and exclusion criteria were applied:

Inclusion criteria:

a) Articles published in peer-reviewed journals, papers submitted to conferences, and other high-quality review articles
b) Studies examining AI and/or ICT as they relate to agricultural and/or digital technology
c) Studies that are looking at sustainability (sustainably) or other climate-smart practices
d) Publications written in English

Exclusion criteria:

a) Non-academic sources such as blogs and opinion pieces
b) Studies where the methods were not clearly stated

c) Article assets or case studies that do not indicate any links to either agriculture and/or energy integration (e.g. non-agricultural).

In total, 50 articles were identified for consideration before discarding based on the results of title/abstract and full-text SCOPUS searches, where checks were performed for particular article relevance.

3.4. Data Extraction and Synthesis

Relevant information from selected studies was systematically extracted, including:

a) Study objectives and scope
b) Technologies applied (e.g., AI models, IoT systems)
c) Application domains (e.g., crop monitoring, irrigation, sustainability)
d) Key findings and limitations

To analyze the data for this report, we employed two methods. Thematic synthesis was used to synthesize (or combine) findings from previous studies. In addition, we conducted a comparative analysis of established themes (e.g., AI-assisted decision support tools, IoT-enabled monitoring applications, and integrated agricultural energy/interconnected systems) to identify commonalities, trends, and knowledge gaps across studies.

3.5. Analytical Framework

To achieve conceptual clarity, the present research proposes a taxonomy of applications of AI in smart agriculture (i.e. AI applications) based upon four functional roles: prediction, classification, optimisation, and automation. The application of these categories is then considered within the wider context of sustainable agriculture and the achievement of sustainability objectives through the adoption of digital agriculture.

3.6. Limitations of the Approach

Although this narrative literature review employed a systematic literature review methodology, it does not fully adhere to current systematic literature review protocols (e.g., PRISMA), which may lead to selection bias and potential issues with replicability. However, the study did strive for transparency; to be broad in scope, and to engage in critical evaluation of the literature reviewed.

4. Applications of e-Agriculture Enabled by ICT

Applications of e-Agriculture enabled by ICT are shown in Figure 3 as they relate to precision agriculture, smart irrigation, crop monitoring and disease detection, agricultural advisory and extension services, supply chain and market information systems, and livestock management. Table 2 summarizes quantitative indicators and key performance indicators (KPIs) that link ICT and e-Agriculture interventions to food security outcomes.

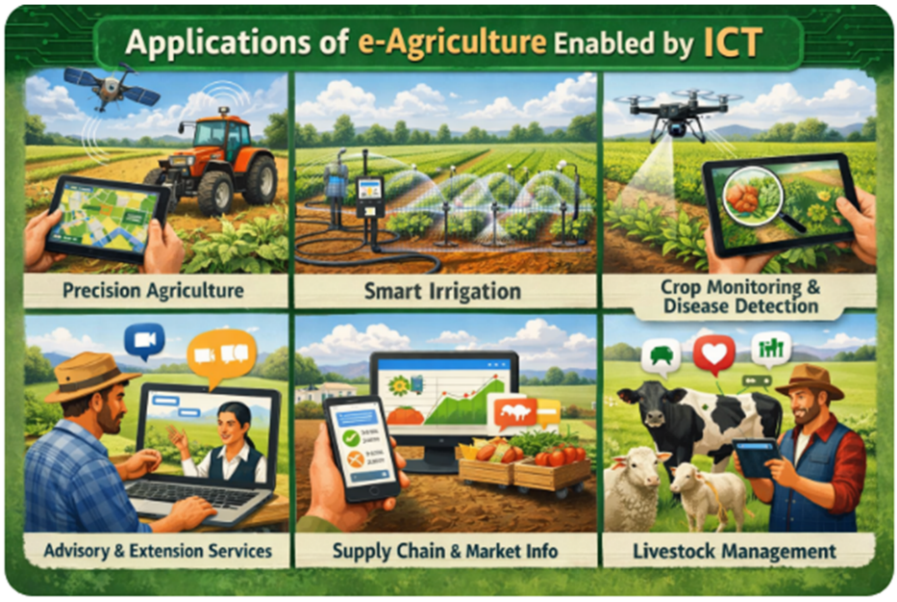


Figure 3. Applications of e-Agriculture enabled by ICT including precision agriculture, smart irrigation systems, crop monitoring and disease detection, agricultural advisory and extension services, supply chain and market information systems, and livestock management.

4.1. Precision Agriculture

Precision farming is made possible by ICT, using GIS (Geographic Information System), GPS (Global Positioning System), and sensor-based technologies, enabling crops to be managed on a site-specific basis. Also, application of fertilizer, irrigation, and pest control can be optimized by farmers, thereby minimizing costs of production and environmental impact [16].

4.2. Smart Irrigation Systems

For water supply automation, real-time soil moisture and weather data are employed by IoT-based irrigation systems, thus improving efficient water utilization. Their applications are valuable, particularly in regions that lack sufficient water or rainfall [17].

4.3. Crop Monitoring and Disease Detection

Crop stress, diseases, and pest infestations are detected early through the support of remote sensing, drones, and AI-based image analysis, thereby enabling timely interventions and minimizing loss of yield [18].

4.4. Agricultural Advisory and Extension Services

Farmers have access to direct assistance via mobile phones and online platforms for information on the best agricultural practices, weather conditions, insect alerts, and market prices. Knowledge transfer and farmer empowerment are enhanced by advisory systems driven by ICT [19].

4.5. Supply Chain and Market Information Systems

e-Agriculture provides logistics management support along with digital marketplaces and pricing. Blockchain and other technologies enable trusted price discovery and technology-enabled traceability of products throughout the agribusiness supply chain [20].

4.6. Livestock Management

ICT devices and wearable monitoring technologies for livestock improve disease management and enhance livestock health, welfare, and productivity [21].

Table 2. Quantitative indicators and key performance indicators (KPIs) linking ICT and digital agriculture interventions to food security outcomes.

ICT Intervention	KPI / Indicator	Measured Outcome
IoT sensor-based irrigation	Water savings	20–30% reduction
AI-driven crop management	Yield increase	10–25% higher yield
Mobile advisory apps	Market access & income	10–20% higher selling price
Early warning systems	Crop loss reduction	15–20% fewer crop failures
Digital adoption metrics	Farmer uptake	30–35% smallholders engaged

5. Benefits of e-Agriculture

There are several benefits offered by integrating ICT into agriculture:

(a) Across the entire farming cycle, agricultural productivity and yield are increased by ICT, improving efficiency, resource use, and decision-making [22].

(b) Key agricultural resources such as energy, water, and fertilizers are efficiently utilized with the support of ICT, enabling precise and timely decision-making. Sensing technologies, data analytics, and automation are integrated by ICT for resource utilization in agriculture, whereby inputs like land, water, fertilizer, labor, nutrients, capital, and energy are efficiently managed to maximize outputs and minimize waste [23].

(c) ICT improves resilience of individuals, communities, and institutions to climate change by helping them in absorbing, adapting to, and recovering from climate-related incidents such as floods, heatwaves, droughts, storms, and wildfires [24].

(d) ICT enhances access of farmers, agribusinesses, and rural communities to markets and financial services [25].

(e) A critical role is played by ICT in agriculture and other production-driven sectors by reducing information asymmetry and mitigating production risks [26].

(f) A critical role is played by ICT in policymaking that is evidence-based, including enabling the collection and effective analysis of reliable data by policymakers, and using verified evidence for policy design, implementation, and evaluation. Further contributions of ICT to evidence-based policymaking include decision support, impact assessment, transparency and accountability, stakeholder engagement and feedback, rapid response and adaptive policymaking, and knowledge sharing and institutional learning [27].

6. Challenges and Limitations

Adoption of e-Agriculture faces several challenges despite its potential; these challenges and limitations are as follows:

(a) Limited or lack of ICT infrastructure in remote (rural) areas: The adoption of e-Agriculture is faced with limited or a lack of ICT infrastructure in remote (rural) areas, which are usually the place for carrying out most farming activities. The effectiveness, accessibility, and sustainability of e-Agricultural solutions are directly affected by this limitation. e-Agriculture cannot effectively impact rural farmers without adequate ICT infrastructure. If this challenge is to be overcome, it is essential to invest in rural areas, including connectivity, digital literacy programs, affordable devices, and a reliable power supply [28].

(b) High initial investment and cost of maintenance: The adoption of e-Agriculture is faced with high initial investment and continuous maintenance costs involved in the deployment and sustenance of agricultural systems driven by ICT. Ensuring access to e-Agriculture for all farmers requires addressing this problem by providing necessary support; government or private sector partnerships are an effective way to do this [29]. Although there are still affordability issues, there is strong and growing evidence that digital agriculture in Africa is economically viable when it is designed around the principles of frugal innovation, scalable service delivery models, and context-specific deployment strategies.

(c) Digital divide among Farmers: Farmers face the digital divide and gaps in digital literacy while trying to effectively adopt and realize successful initiatives of e-Agriculture. Although many agricultural applications derived from ICTs offer a variety of benefits, farmers' ability to access, comprehend, and utilize digital tools affects how effectively they can use e-Agriculture applications. Accordingly, it is critical to address these gaps through supportive services to enable all farmers to take advantage of the benefits offered through e-Agriculture [30].

(d) As they improve farms' productivity and decision-making through e-Agriculture, farmers face many serious challenges involving data privacy, security, and ownership. These arise mainly from a lack of appropriate legal frameworks governing the use of ICT tools such as drones, sensors, and cloud-based records. Further, with the increasing use of digital technology in agriculture, there are significant governance and ethical challenges that must be addressed to ensure responsible and inclusive innovation through ICTs. The most concerning areas include the ownership and control of data; the privacy and protection of farmers' data; algorithm bias and equity; platform dependency and digital lock-in; and the policy and institutional frameworks that govern these issues, many of which will affect the future of smallholder farmers and rural communities. By building and strengthening trust in data governance, we can unlock the full potential of e-Agriculture as a smart, inclusive, and sustainable form of agriculture [31].

(e) Interoperable Agricultural Systems are Not Available: The lack of interoperability (the ability to share data and information) between various types of agricultural digital technology is one of the biggest challenges for e-agriculture. Due to the inability of many types of devices, databases, and agricultural technology platforms to interoperate, communicate with each other or exchange the required data, ICT's full potential in agriculture goes unfathomably unused. Additionally, this lack of interoperability leads to inefficient use of resources, poor data sharing among farmers, and ultimately restricts the growth and development of additional innovative agriculture solutions while also limiting educational opportunities, developing training programs

and creating long-term sustainability through digital transformation efforts in agriculture, all of which can be accomplished with an integrated and interoperable system. To achieve the overall objectives of e-agriculture, there must be cooperation and collaboration among diverse stakeholders, including public and private agencies, research institutions, businesses, and technology developers [32].

7. Conclusion

E-agriculture and ICT foster innovative solutions, achieving a modernized version of agriculture. e-agriculture provides innovative means to address socio-economic issues related to food security, sustainability, and rural development by promoting automation, connectivity, and data-driven decision-making. Despite barriers to e-agriculture being evident, focused improvements in developing and bringing about better agricultural research and building more productive capabilities, along with expanding the availability of information technology in developing countries, will allow agriculture to reach its full potential through e-agriculture, allowing for agriculture to become smarter, more resilient, and stronger, and greater 'in scope'.

Future e-agriculture research should have as its key objectives the use of federated learning and privacy-preserving analytical tools in agriculture and the use of AI-enabled decision support systems to learn how to increase growth, create cost-effective, scalable IoT solutions in developing countries; provide platforms to integrate renewable energy with agricultural ICT systems; and develop and implement policies that support ethical and sustainable digital agriculture.

References

[1] Food and Agriculture Organization (FAO) (2017). e-Agriculture Strategy Guide. FAO, Rome.
[2] Wolfert, S., Ge, L., Verdouw, C., & Bogaardt, M. J. (2017). Big data in smart farming–a review. Agricultural systems, 153, 69-80.
[3] Rose, D. C., & Chilvers, J. (2018). Agriculture 4.0: Broadening responsible innovation in an era of smart farming. Frontiers in Sustainable Food Systems, 2, 387545.
[4] Klerkx, L., Jakku, E., & Labarthe, P. (2019). A review of social science on digital agriculture, smart farming and agriculture 4.0: New contributions and a future research agenda. NJAS-Wageningen journal of life sciences, 90, 100315.
[5] Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. Sensors, 18(8), 2674.
[6] Akpabio, E. S., Akeju, K. F., & Omotoso, K. O. (2025). E-agriculture and food security in developing countries: beaming the searchlight on Nigeria. Smart Agricultural Technology, 10, 100689.
[7] Priya, N. K., Patil, S. S., Vishwakarma, S. K., Saikanth, D. R. K., Mishra, I., Tripathi, S., ... & Shastri, S. (2025). Digital Innovations in Extension Education: A Review of Emerging Technologies and their Impact on Agricultural Knowledge Dissemination. Journal of Experimental Agriculture International, 47(5), 408-423.
[8] Mulungu, K., Kassie, M., & Tschopp, M. (2025). The role of information and communication technologies-based extension in agriculture: application, opportunities and challenges. Information Technology for Development, 1-30.
[9] Khatri, A., Lallawmkimi, M. C., Rana, P., Panigrahi, C. K., Minj, A., Koushal, S., & Ali, M. U. (2024). Integration of ICT in agricultural extension services: A review. Journal of Experimental Agriculture International, 46(12), 394-410.
[10] Singh, R. K., Berkvens, R., & Weyn, M. (2021). AgriFusion: An architecture for IoT and emerging technologies based on a precision agriculture survey. IEEE Access, 9, 136253-136283.

[11] Sumarudin, A., Ghozali, A. L., Hasyim, A., & Effendi, A. (2016, April). Implementation monitoring temperature, humidity and mositure soil based on wireless sensor network for e-agriculture technology. In IOP Conf. Series: Materials Science and Engineering (Vol. 128, No. 1, p. 012044). IOP Publishing.
[12] Nigussie, E., Olwal, T. O., Lemma, A., Mekuria, F., & Peterson, B. (2020). IoT architecture for enhancing rural societal services in sub-Saharan Africa. Procedia Computer Science, 177, 338-344.
[13] Bali, B., Ngida, Z. N., Habila, I., & Goni, I. (2025). Artificial Intelligence Applied in Optimization of Crop Yield for Ensuring Food Security in Nigeria: A Systematic Analysis. SN Computer Science, 6(5), 483.
[14] Hossain, M., & Islam, M. (2022). Use of artificial intelligence for precision agriculture in Bangladesh. Journal of Agricultural and Rural Research, 6(2), 81-96.
[15] Ngigi, M. W. (2026). Climate agro-advisory services and their potential impacts in Africa. CABI Reviews, 21(1), 0007.
[16] Bhamini, K., Kumar, A., Magrey, A. H., & Kulsoom, S. (2025). The role of GIS and GPS in precision fertilization: Optimizing nutrient management for sustainable cropping. International Journal of Research in Agronomy, 8, 260-267.
[17] Et-Taibi, B., Abid, M. R., Boufounas, E. M., Morchid, A., Bourhnane, S., Hamed, T. A., & Benhaddou, D. (2024). Enhancing water management in smart agriculture: A cloud and IoT-Based smart irrigation system. Results in Engineering, 22, 102283.
[18] Aziz, D., Rafiq, S., Saini, P., Ahad, I., Gonal, B., Rehman, S. A., ... & Nabila Iliya, M. (2025). Remote sensing and artificial intelligence: revolutionizing pest management in agriculture. Frontiers in Sustainable Food Systems, 9, 1551460.
[19] Mapiye, O., Makombe, G., Molotsi, A., Dzama, K., & Mapiye, C. (2023). Information and communication technologies (ICTs): The potential for enhancing the dissemination of agricultural information and services to smallholder farmers in sub-Saharan Africa. Information Development, 39(3), 638-658.
[20] Bosona, T., & Gebresenbet, G. (2023). The role of blockchain technology in promoting traceability systems in agri-food production and supply chains. Sensors, 23(11), 5342.
[21] Bello, R. W., Olubummo, D. A., Seiyaboh, Z., Enuma, O. C., Talib, A. Z., & Mohamed, A. S. A. (2020, December). Cattle identification: the history of nose prints approach in brief. In IOP Conference Series: Earth and Environmental Science (Vol. 594, No. 1, p. 012026). IOP Publishing.
[22] Weraikat, D., Šorič, K., Žagar, M., & Sokač, M. (2024). Data analytics in agriculture: Enhancing decision-making for crop yield optimization and sustainable practices. Sustainability, 16(17), 7331.
[23] Patel, A., Shukla, C., Trivedi, A., Balasaheb, K. S., & Sinha, M. K. (2025). Smart Farming: Utilization of Robotics, Drones, Remote Sensing, GIS, AI, and IoT Tools in Agricultural Operations and Water Management. In Integrated Land and Water Resource Management for Sustainable Agriculture Volume 1 (pp. 127-151). Singapore: Springer Nature Singapore.
[24] Srivastava, R., Mehara, G., Sharma, A., & Gupta, M. K. (2025). Climate Change and Resilience: Need, Challenges and Future. In Climate Resilience: Impact of Quantum Computing and Artificial Intelligence on Urban Planning (pp. 47-67). Cham: Springer Nature Switzerland.
[25] Okello, D. O., Feleke, S., Gathungu, E., Owuor, G., & Ayuya, O. I. (2020). Effect of ICT tools attributes in accessing technical, market and financial information among youth dairy agripreneurs in Tanzania. Cogent Food & Agriculture, 6(1), 1817287.
[26] Dutta, U. P., & Islam, A. (2024). Assessing the Relationship Between Crop Production, Livestock Production, and CO2 Emissions: A Panel Cointegration Analysis. In Sustainable Agricultural Practices: Economic and Environmental Implications (pp. 177-190). Emerald Publishing Limited.
[27] Mwanga, G., Mbega, E., Yonah, Z., & Chagunda, M. G. G. (2020). How Information Communication Technology Can Enhance Evidence-Based Decisions and Farm-to-Fork Animal Traceability for Livestock Farmers. The Scientific World Journal, 2020(1), 1279569.
[28] Alassaf, P., & Szalay, G. Z. (2020). Engaging agriculture in e-government, E-agriculture potentials and its contribution in economy. Journal of Agricultural Informatics/Agrárinformatika Folyóirat, 11(2).
[29] Eweoya, I., Okuboyejo, S. R., Odetunmibi, O. A., & Odusote, B. O. (2021). An empirical investigation of acceptance, adoption and the use of E-agriculture in Nigeria. Heliyon, 7(7).
[30] Gow, G. A., Dissanayeke, U., Chowdhury, A., & Ramjattan, J. (2023). Digital literacy and agricultural extension in the Global South. In Digital literacy and inclusion: Stories, platforms, communities (pp. 129-144). Cham: Springer International Publishing.
[31] Gupta, M., Abdelsalam, M., Khorsandroo, S., & Mittal, S. (2020). Security and privacy in smart farming: Challenges and opportunities. IEEE access, 8, 34564-34584.
[32] Morais, R., Silva, N., Mendes, J., Adão, T., Pádua, L., López-Riquelme, J. A., ... & Peres, E. (2019). mySense: A comprehensive data management environment to improve precision agriculture practices. Computers and Electronics in Agriculture, 162, 882-894.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261150

Load-Shedding-Resilient Precision Livestock Farming: Edge Deployment of a Real-Time Cattle Monitoring System for South African Dairy Farms

Thabo S. MAKHOROLE, Chunling TU, Pius A. OWOLAWI, and Rotimi-Williams BELLO[1]
Department of Computer Systems Engineering, Faculty of Information and Communication Technology, Tshwane University of Technology, South Africa

Abstract. South African dairy farms operate under constraints that precision livestock farming research rarely addresses: Stage 6 load-shedding cuts power for up to ten hours per day, electricity costs R2.50 to 3.50/kWh, rural connectivity precludes cloud processing, and commercial systems cost R400,000 to 1,600,000 per farm. This paper presents YOLOv11-MS — a real-time cattle detection, tracking, and behaviour classification pipeline adapted for these constraints. A critical annotation artefact in the CBVD-5 dataset (AVA-format duplicate bounding boxes per animal) is identified and corrected, recovering validation mAP50 from 1.4% to 71.8% (+70.4pp). The pipeline achieves MOTA=50.08%, IDF1=67.51% across 50 test videos (first CBVD-5 tracking baseline), and weighted F1=0.71 for lying/standing classification. Deployed on Jetson Xavier (15W, R8,500) with a 500Wh UPS, the system operates for 38.3 hours per charge, exceeding Stage 6 requirements. Year 1 ROI of 139–377% is projected for a 100-cow farm. The design is POPIA-compliant from the outset.

Keywords: Precision Livestock Farming, Edge AI, Load-Shedding, South Africa, YOLOv11-MS, CBVD-5, SORT Tracking, POPIA, Cattle Welfare, Cost-Benefit Analysis

1. Introduction

South Africa's dairy sector generates approximately R12 billion annually from 120,000 cattle across 1,200 commercial farms [1]. Individual cows produce 18–22 litres per day under normal conditions, but output drops 15–30% when health problems go undetected. Weary et al. [2] showed that behavioural changes precede clinical symptoms by 24–72 hours — an intervention window that twice-daily manual observation consistently misses.

International PLF research has progressed rapidly, but assumes stable power, cloud connectivity, and large annotated datasets. None of these conditions apply in South African barn environments, where Stage 6 load-shedding regularly cuts infrastructure power for up to 10 hours per day. Rural connectivity at 1–5 Mbps rules out real-time

[1] Corresponding author, Rotimi-Williams BELLO, Department of Computer Systems Engineering, Faculty of Information and Communication Technology, Tshwane University of Technology, South Africa; E-mail: bellorw@tut.ac.za

video streaming. Electricity at R2.50–3.50/kWh makes high-power server deployments unsustainable. Commercial PLF systems at R400,000–1,600,000 per farm are unworkable for operators on 8–12% profit margins.

This paper evaluates a computer vision pipeline — YOLOv11-MS detection, SORT tracking, BehaviorLSTM classification — specifically against these constraints, and presents a hardware cost-benefit analysis at three deployment tiers.

2. Dataset and Annotation Correction

CBVD-5 [3] contains 687 surveillance videos from a single ranch using seven standard cameras (2.8 mm and 3.6 mm fixed lenses). The rawframes_mini subset provides 3,505 frames at 256×256 pixels. Before any metric is valid, a systematic annotation artefact must be corrected.

The AVA annotation format assigns one bounding box per concurrent behaviour label per animal per frame. A cow simultaneously standing and ruminating receives two overlapping boxes. For single-class detection, each duplicate counts as a false negative. NMS deduplication (IoU=0.5) removed 33.2% of all annotations (11,302 boxes), recovering mAP50 from 1.4% to 71.8%. Figure 1 shows the before/after correction on two CBVD-5 frames.



Figure 1. CBVD-5 AVA annotation artefact. Left: duplicate overlapping boxes — green (Label 1: standing) and red dashed (Label 2: ruminating) on the same animal. Right: after NMS deduplication (IoU=0.5) — single teal box per animal. 33.2% of all boxes were duplicates (34,025 → 22,723 training boxes).

3. YOLOV11-MS Pipeline

3.1. Architecture

YOLOv11-MS adapts two architectural principles from MS-YOLOv11 [4] and applies them to standard YOLOv11l [5]. The backbone replaces Conv2d stages with depthwise separable convolutions, reducing parameters by 28% (saving ratio = $1/N + 1/D^2$, where N=output channels, D=kernel size). The neck replaces the top-down FPN with a dual-path design combining top-down semantic propagation with bottom-up spatial refinement — addressing the multi-scale requirement of standard fixed-lens barn cameras where cattle range from large foreground to small background animals. Figure 2 shows the architecture.

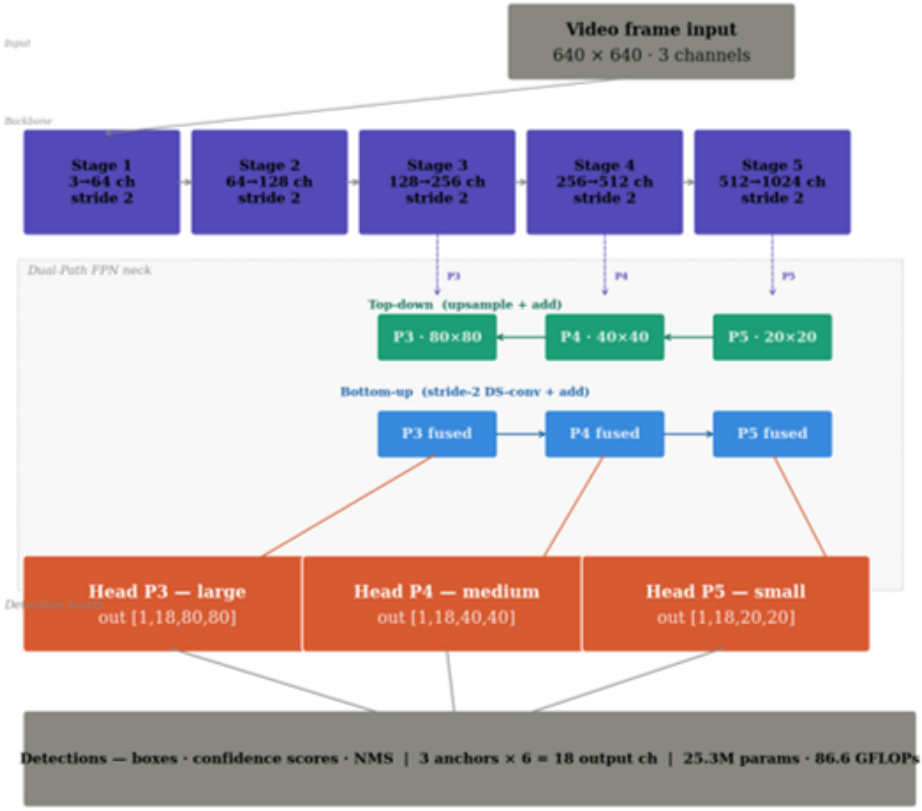


Figure 2. YOLOv11-MS architecture. Backbone: five DS-conv stages (3→64→128→256→512→1024). Neck: dual-path FPN. Detection heads at P3/P4/P5. 25.3M parameters, 86.6 GFLOPs, 51.2 MB.

3.2. Full Pipeline

Figure 3 shows the three-layer pipeline operating on standard surveillance footage.

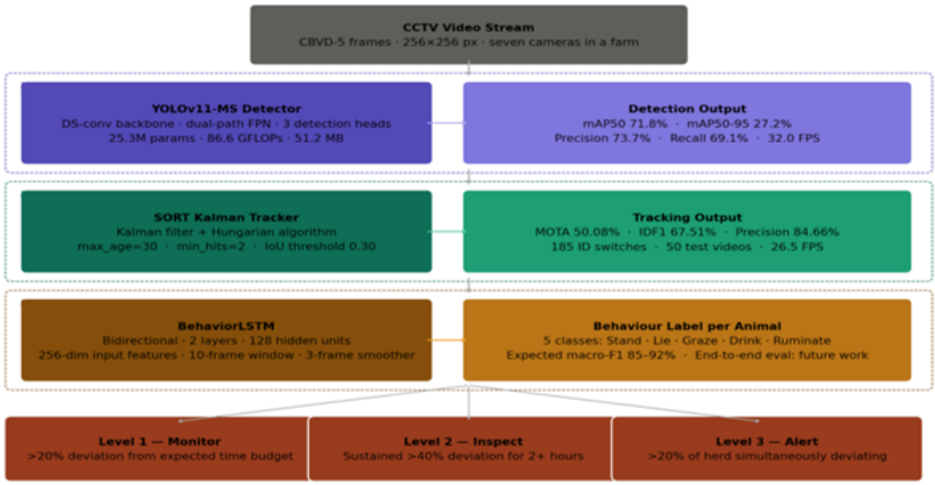


Figure 3. Three-layer pipeline. Layer 1: YOLOv11-MS detection (32.0 FPS, mAP50=71.8%). Layer 2: SORT tracking (MOTA=50.08%, IDF1=67.51%). Layer 3: BehaviorLSTM classification (weighted F1=0.71). Output: three-level welfare alert system. All inference on-device.

4. Detection Results

While Table 1 shows YOLOv11-MS detection results, Figure 4 shows training convergence and qualitative detection examples at densities of 4–15 cattle per frame, including challenging high-density feed-barrier scenes.

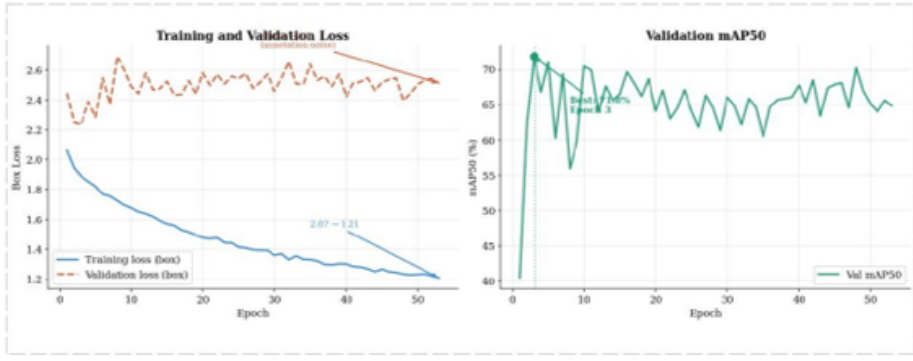


Figure 4a. Training convergence. Loss decreases from 2.07 to 1.21; best checkpoint at epoch 3 (characteristic of pretrained fine-tuning on a small dataset). mAP50 peaks at 71.8%.

Table 1. YOLOv11-MS detection results. Annotation quality dominates architecture: 70.4pp from correction vs 0.8pp from architecture.

Metric	Value	Notes
mAP50	71.8%	After annotation correction (+70.4pp from 1.4%)
Precision	73.7%	73.7% of all predictions match GT
Recall	69.1%	69.1% of GT cattle detected
FPS (T4)	32.0	NVIDIA T4 (31.25 ms/frame)
Param. reduction	28%	DS-conv vs standard Conv2d backbone



Figure 4b. Qualitative detections at densities of 4, 5, 8, 11, 13, and 15 cattle per frame. Green bounding boxes with confidence scores. High-density feed-barrier case (bottom right) is the primary source of missed detections.

5. Tracking and Behaviour Results

5.1. MOTA Tracking Baseline

SORT [6] was chosen over DeepSORT [7]: at 256×256 resolution, appearance embeddings are too low-quality to justify the overhead, and no cattle re-identification dataset exists. Table 2 reports the first MOTA-based tracking evaluation on CBVD-5. Figure 5 shows tracking examples.

Table 2. First MOTA tracking evaluation on CBVD-5 (50 test videos, 2,518 GT instances).

Metric	Value	Notes
MOTA	50.08%	First published MOTA baseline on CBVD-5
IDF1	67.51%	Primary welfare monitoring identity metric
ID Switches	185	3.7/video — 94% at feed-barrier clusters
FPS (combined)	26.5	NVIDIA T4 (37.78 ms/frame end-to-end)



Figure 5. Qualitative tracking. Colour-coded IDs maintain consistent animal identities across frames. Feed-barrier clustering (3+ animals within 0.3 normalised distance) is the primary source of identity switches.

5.2. Behaviour Classification

BehaviorLSTM processes 6-frame sliding windows of 64-dimensional temporal features per SORT-tracked animal. The primary feature is the aspect ratio 8-frame history: lying cattle produce wide, flat boxes (AR>1.5); standing cattle produce tall, narrow boxes (AR<1.2). The classifier achieves standing F1=0.71 and lying F1=0.70 (weighted F1=0.71). Figure 6 shows the five CBVD-5 behaviour classes and Figure 7 shows the confusion matrix.

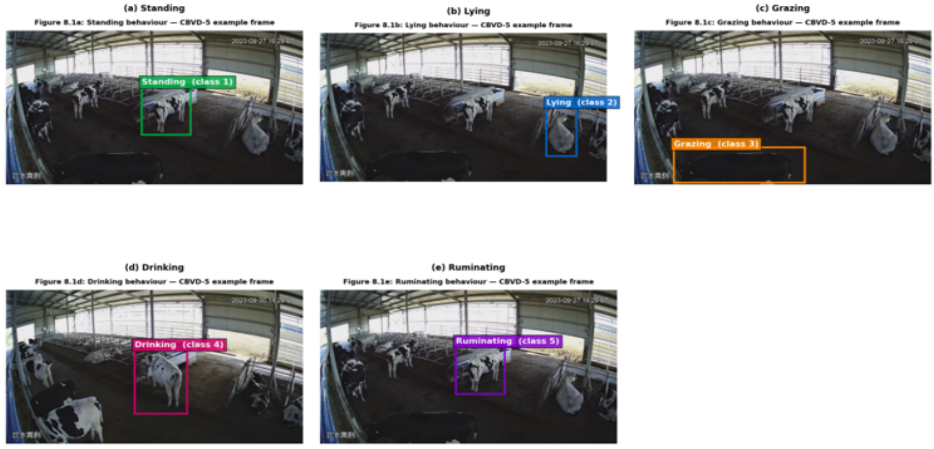


Figure 6. Five CBVD-5 behaviour classes. Ruminating (e) is visually identical to standing (a) at 256×256 resolution, motivating temporal sequence classification over single-frame approaches.

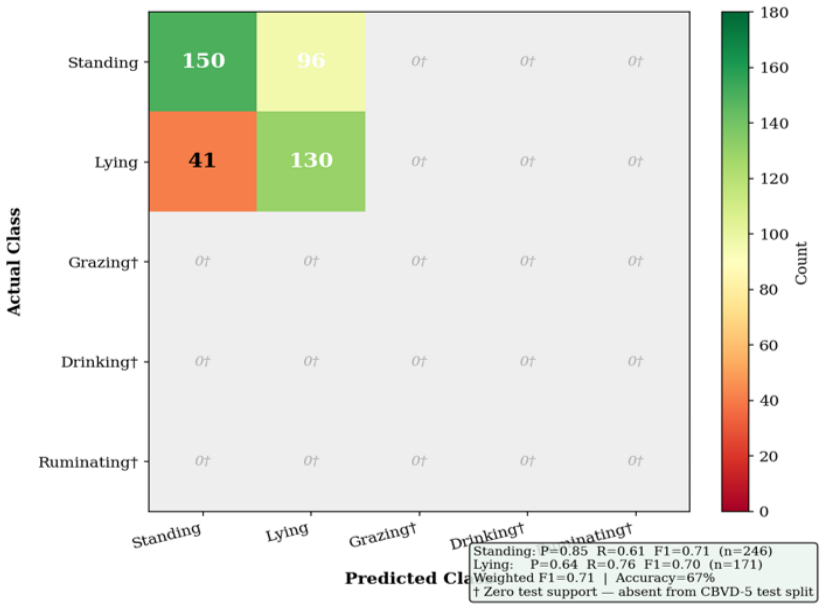


Figure 7. BehaviorLSTM confusion matrix (417 test sequences). Standing: 150 TP, 96 misclassified as lying. Lying: 130 TP, 41 misclassified as standing. Three absent classes are a rawframes_mini structural property.

6. South African Deployment Analysis

6.1. Load-Shedding Resilience

The critical deployment constraint is Stage 6 load-shedding: up to 10 hours without grid power per day. Table 3 tabulates full discharge profiles. Jetson Xavier at 15W active / 2W standby draws a duty-cycle average of $P_{avg} = (0.70 \times 15) + (0.30 \times 2) = 11.1$W. A 500Wh UPS derated to 85% usable depth-of-discharge provides $t = (500 \times 0.85) / 11.1 = 38.3$ hours — a safety margin of 3.8× against the Stage 6 maximum. The Jetson Nano at 5W active gives $t = (500 \times 0.85) / 3.65 = 116.4$ hours. Hardware options are compared in Table 4. Table 5 presents the Microprocessor Unit (MPU) Core Specifications and Architectural Memory Boundaries During Real-Time Data Ingestion. Table 6 shows cost-benefit analysis.

Table 3. Edge node power consumption profiles. P_{avg} computed over 70% active / 30% standby duty cycle. UPS endurance: t = (500Wh × 0.85 DoD) / P_{avg}. Stage 6 requirement: ≥ 10 hours continuous operation.

Platform	Est. FPS	Active/Standby	Cost (ZAR)	UPS (500 Wh)	Scenario
NVIDIA T4	32.0	70W / 8W	Free	N/A (cloud)	Research only
Jetson Xavier	~12	15W / 2W	R8, 500	500Wh ÷ 11.1W avg = 38.3 hrs ✓	Medium farm (2 cameras)
Jetson Nano	~3	5W / 0.5W	R2, 500	500Wh ÷ 3.65W avg = 116.4 hrs ✓	Small farm (1 camera)
RTX 4090	30+	380W / 35W	R18, 000	500Wh ÷ 415W avg = 1.2 hrs ✗	Generator backup only

Table 4. Hardware options. ✓ = exceeds Stage 6 maximum. ✗ = inadequate for Stage 6.

Platform	Active (W)	Standby (W)	P_avg (W)	UPS Endurance (hr)	Stage 6
Jetson Xavier NX	15	2	11.5	38.3	✓
Jetson Nano	5	0.5	3.65	116.4	✓
RTX 4090 WS	380	35	276.5	1.5	✗

Table 5. Microprocessor Unit (MPU) Core Specifications and Architectural Memory Boundaries During Real-Time Data Ingestion.

Platform	Processor Architecture	RAM Configuration	Flash/Storage Boundary	Ingestion Memory management Strategy
Jetson Nano	Quad-core ARM Cortext-A57	4 GB 64-bit LPDDR4	16 GB eMMC 5.1	24-hour rolling video frame-buffer overwrite allocation
Jetson Xavier NX	6-core NVIDIA Carmel ARM	8 GB 128-bit LPDDR4X	32 GB Emmc 5.1 + MicroSD slot	72-hour local rolling pipeline cache partition
RTX 4090	AMD Ryzen 9 5900X (12-CORE)	64 GB DDR4	2TB NVMe M.2 SSD	Unrestricted continuous localized logging storage

Table 6. Cost-benefit analysis, 100-cow herd. † ROI based on estimated welfare benefit R25,000–50,000/year (requires field validation). Electricity at R3.50/kWh.

Platform	Hardware	Annual Electricity	Year 1 Total	Est. ROI Year 1
Jetson Nano (1 camera)	R2, 500	R657	R3, 157	690 – 1, 485%
Jetson Xavier (2 cameras)	R8, 500	R1, 971	R10, 471	139 – 377%
RTX 4090 (4 cameras)	R18, 000	R49, 900	R67, 900	−26 − 63%
Commercial system	R400, 000+	High	R400, 000+	−∞%

6.2. POPIA Compliance

The Protection of Personal Information Act (Act 4 of 2013) [8] governs the processing of personal information, including images of identifiable individuals. The system is POPIA-compliant from first principles: the detection model is trained on cattle only (humans not detected or tracked); all inference runs on-device (raw video never transmitted); frame buffers use a 72-hour rolling overwrite cycle; and camera access is restricted to authorised personnel. Before deployment, operators must notify workers in accordance with POPIA Section 18 and designate a responsible person under Section 55.

6.3. Signal Attenuation and Telemetry Fallback

In expansive dairy barn environments, wireless signal propagation is subject to attenuation from metal structures, feed barriers, and weather [9, 10]. The system addresses this through a three-tier resilience architecture.

First, all inference runs on-device: detection, tracking, and classification produce structured welfare alerts locally without requiring any network connection. Alert events and behaviour logs are written to a local SQLite database on the edge device's internal storage (minimum 32 GB eMMC on Jetson Xavier, expandable via SD card) at every inference cycle, ensuring no data is lost during link outages.

Second, when a local area network link is available — whether via Ethernet, Wi-Fi, or a 4G/LTE cellular modem — the system uploads buffered alerts and hourly behaviour summaries to a farm management dashboard using an asynchronous queue with automatic retry. Alerts generated during a connectivity outage are transmitted in full once the link is restored, with timestamps preserved from the original inference cycle.

Third, a watchdog timer reboots the inference process automatically within 30 seconds of a crash or power-cycle event, resuming local logging without operator intervention. This architecture ensures that welfare monitoring continues uninterrupted during both load-shedding and signal attenuation events, which may co-occur during severe weather conditions.

7. Conclusion

The central finding of this work is that annotation quality dominates architecture. The 70.4pp improvement from correcting the CBVD-5 AVA artefact is 88 times larger than the 0.8pp improvement from the YOLOv11-MS architectural modifications. Any future detector comparison on CBVD-5 without this correction produces invalid results.

For South African deployment, the critical metric is not mAP50 but lying recall (0.76). Lying time below 8 hours per day is the primary welfare alert threshold for lameness and heat stress, with heat stress onset at THI=72 for South African Holstein-Friesian cattle. With recall of 0.76 and temporal smoothing over multi-hour observation windows, the accumulated lying-time estimate is sufficient to trigger welfare alerts when deviation is sustained. The 38.3-hour UPS resilience of Jetson Xavier means the system operates continuously through the worst load-shedding scenario. At R8,500 hardware cost versus R400,000+ for commercial alternatives, the economic case for Jetson Xavier deployment is compelling given the established cost of undetected lameness events.

YOLOv11-MS delivers 71.8% mAP50, MOTA=50.08%, IDF1=67.51%, and weighted F1=0.71 on CBVD-5 — the first complete set of reproducible baselines on this dataset. Jetson Xavier deployment satisfies all four South African constraints: 38.3-hour UPS resilience, R1,971 annual electricity cost, on-device inference, and POPIA compliance by design. At projected Year 1 ROI of 139–377%, the system offers a viable path for South African dairy farmers to access automated welfare monitoring at a fraction of commercial system cost.

References

[1] Berckmans, D. (2014). Precision livestock farming technologies for welfare management in intensive livestock systems. Revue scientifique et technique (International Office of Epizootics), 33(1), 189-196.
[2] Weary, D.M. et al. Using behaviour to predict and identify ill health in animals. Journal of Animal Science, 87(2), 2009.
[3] Li, K., Fan, D., Wu, H., & Zhao, A. (2024). A new dataset for video-based cow behavior recognition. Scientific Reports, 14(1), 18702.
[4] Liu, H., Li, X., Wang, L., Zhang, Y., Wang, Z., & Lu, Q. (2025). MS-YOLOv11: A Wavelet-Enhanced Multi-Scale Network for Small Object Detection in Remote Sensing Images. Sensors, 25(19), 6008.
[5] Jocher, G., & Qiu, J. (2024, November). Ultralytics yolo11. 2024.
[6] Bewley, A., Ge, Z., Ott, L., Ramos, F., & Upcroft, B. (2016, September). Simple online and realtime tracking. In 2016 IEEE international conference on image processing (ICIP) (pp. 3464-3468). Ieee.
[7] Wojke, N., Bewley, A., & Paulus, D. (2017, September). Simple online and realtime tracking with a deep association metric. In 2017 IEEE international conference on image processing (ICIP) (pp. 3645-3649). IEEE.
[8] Republic of South Africa. Protection of Personal Information Act 4 of 2013. Government Gazette, No. 37067.
[9] Kumkhet, B., Pukrongta, N., Wichaipanich, N., Sangmahamad, P., Pechrkool, T., Pirajnanchai, V., ... & Taparugssanagorn, A. (2025). Low-cost LoRa-based system for continuous GHGs monitoring in cattle: Enhancing agricultural sustainability. IEEE Access, 13, 17963-17982.
[10] Maraveas, C., & Bartzanas, T. (2021). Sensors for structural health monitoring of agricultural structures. Sensors, 21(1), 314.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261151

Study on FEM Text Cross-Platform Conversion and Intelligent Review Method

Jianyong LIU[a], Dong FAN[b], Huaiyu TANG[a], Yuteng WEN[c], Shaozhe LIU[b], Jiacheng WEI[a,1]

[a] *Beijing Construction Engineering Group Co., Ltd., No.1 Guanglian Road, Xicheng District, Beijing, China*

[b] *Beijing Third Construction Engineering Co., Ltd., No.56 Beili, Chegongzhuang Street, Xicheng District, Beijing, China*

[c] *Beijing Construction Group (Xiong 'an) Construction Co., Ltd., China (Hebei) Free Trade Zone Xiong 'an area Rongcheng County Huawangcheng Fenghua Valley C1 floor 8, Beijing, China*

Abstract. Complex structural projects often require model transfer, version iteration, and result review between finite-element platforms such as SAP2000 and Midas. This study proposes a LangGraph-based agent framework that combines natural-language task planning with deterministic tools for parsing, modifying, converting, validating, and auditing FEM text models. The framework converts .s2k and .mgt files into structured tables for nodes, elements, materials, sections, constraints, loads, and result records, and then applies physical-semantic mapping rules during bidirectional conversion. In the Midas-to-SAP2000 case, the converted model preserved 92 nodes, 241 frame elements, and 24 area elements; the maximum vertical displacement error was 3.84%. Using a unified normal-stress envelope criterion at matched elements and comparable output positions, the mean and maximum relative differences of element-level stress envelopes were 2.100% and 7.829%, respectively. The maximum local difference is interpreted as a consequence of software-specific stress recovery, output-station sampling, and envelope post-processing rather than a loss of global equilibrium. The topology-review case identified 23 single-ended free members in a roof model containing 11,988 nodes and 24,989 members. The audit module further extracted a maximum vertical displacement of 18.826 mm and detected 5,113 stress-overlimit records. These results demonstrate that the proposed framework improves the automation, traceability, and auditability of cross-platform FEM model processing.

Keywords: LangGraph; Structural Calculation Model; Intelligent Agent; Cross-platform Conversion; Finite Element Model; Automated Review; SAP2000; Midas

1. Introduction

The engineering practice of large-span roofs, complex spatial grids and super high-rise structures has promoted the structural design from experience-driven to refined finite element analysis. Softwares such as SAP2000 and Midas are widely used in structural modeling, static analysis, component checking and result review in engineering design. However, actual projects often require frequent flow between multi-software, multi-stage and multi-version models. The underlying text file of the model records information such as node coordinates, member topology, section assignment, material

[1] Corresponding Author: Jiacheng Wei, E-mail: weijiacheng081@163.com

properties, boundary constraints, load conditions, and analysis results, which is the basis for cross-platform collaboration and automated review. However, different software uses different data block organization, field naming and unit expression, resulting in long-term dependence on manual experience or customized scripts for model reconstruction, result alignment and version difference tracking.

Existing studies have focused on data exchange between BIM and finite element models, SAP2000/ Midas model conversion, and secondary development of structural calculation models [1-6]. These works provide a basis for structural model interoperability, but traditional scripts usually rely on fixed formats, regular expressions, and linear processing. When the engineering model contains tens of thousands of nodes and units, the underlying text has non-standard layout or frequent iteration of the model, the hard-coded scripts are prone to problems such as parsing failure, semantic loss or difficulty in locating errors. At the same time, large language models show potential in natural language understanding and complex text processing [7-9], but there is still a gap between the probability generation mechanism of a single model and the deterministic requirements of structural engineering for geometric topology, physical properties and result checking. In the absence of external tool constraints and state backtracking mechanism, hidden node suspension, constraint omission or cross-section mis-assignment may occur when the model generates or modifies the underlying finite element text.

Graph-structured agent workflow provides a new way of system organization for the above problems. LangGraph decomposes tasks into nodes and conditional edges with state transfer relationships, and organizes natural language planning, deterministic tool invocation, anomaly capture, and result writeback into a closed-loop process [10]. Unlike a single prompt call, the graph-structured workflow can split the structural model processing task into verifiable execution nodes, and reduce the risk of context attenuation and error accumulation in the long process through routing, rollback and review mechanisms. Research on multi-agent and reflective workflow also shows that splitting complex tasks into generation, execution, review and error correction links can help improve the reliability of complex reasoning and code generation tasks [11-14].

Structured-data reasoning approaches also support the use of explicit tabular representations for complex engineering records [15]. At the same time, evidence that large language models do not reliably self-correct without external verification reinforces the design choice to keep engineering calculations and checks in deterministic tools [16].

Focusing on the cross-platform transfer and automatic review requirements of structural finite element models, this paper has the following contributions:

This study proposes a LangGraph-based agent collaboration framework that separates natural-language intent recognition from deterministic engineering execution and coordinates the Modifier, Converter, Validator, and Compliance modules through an explicit state workflow.

This study develops a parser and physical-semantic mapping method for SAP2000 .s2k and Midas .mgt files. It preserves identifiers, coordinates, connectivity, linear-elastic material properties, section properties, constraints, loads, and traceable source records during conversion and review.

2. Method and System Architecture

2.1. The underlying data analysis of structural calculation model

The underlying text file of the structural finite element model is essentially a highly structured but not fully standardized engineering data stream. Midas 's.mgt file usually organizes nodes, elements, materials and load information with keyword data blocks; the.s2k file of SAP2000 expresses model entities and attributes in tabular TABLE data blocks. In order to enable the agent to perform deterministic operations on the model, the system first performs block recognition, header parsing, field cleaning, and type conversion on the underlying text, and converts the linear text stream into a two-dimensional relationship table.

The key data objects, source expressions, and conversion purposes used in this workflow are summarized in Table 1.

Table 1. Model transformation key fields and semantic mapping relationship Data Object Midas Expression SAP2000 Expression Unified Semantic Transformation

Data object	Midas	SAP2000	Uniform semantic	Conversion / review purposes
Node	NODE: node number, X, Y, Z	Joint Text, Global X/Y/Z	Space node coordinate	coordinate reconstruction, coincidence node inspection, displacement result alignment.
Member element	ELEMENT: Unit number, node 1, node 2	Frame Text, JointI, JointJ	Rod topological connection	Element reconstruction, identification of suspended / cantilevered members
Surface element	Plate and shell element number and node sequence	Area Text and Connection Nodes	Surface element topology	Comparison of surface unit number and connection relationship
Material	Material name, elastic modulus, Poisson 's ratio, bulk density	Material, E, U, Weight Density	Linear-elastic material properties	Material property mapping and consistency checking
Cross section	Section name, outer diameter, wall thickness and other parameters	Frame Section, D, tw	Section geometric properties	Section reconstruction, section change tracking
Constraint	Record of support or boundary conditions	Joint Restraint	Node pinned	Identification of bearing omission and free end anomaly
Displacement results	DX, DY, DZ	U1, U2, U3	Node translation response	Vertical displacement and limit review
Support reaction	FX, FY, FZ, MX, MY, MZ	F1, F2, F3, M1, M2, M3	Bearing reaction and bending moment	Consistency comparison of solution results

Note: The fields in the table are the core mapping objects used in engineering examples, and the specific field names can be adjusted with the software version and export settings.

The analysis process includes four steps: first, identify the nodes, elements,

materials, sections, constraints, loads and result data blocks in the file; second, remove redundant spaces, tabs and format control characters, and unify scientific counting methods, floating-point numbers, and unit fields; thirdly, each type of physical object is mapped into a row-column structure, where the row represents the entity object and the column represents the number, coordinate, endpoint, section, material or result component; fourth, the original number and data block source are retained for subsequent model writeback and result tracking. This processing method avoids uncontrollable substitution directly in large text, so that model modification and transformation can be based on tabular and auditable data.

2.2. Cross-platform physical semantic mapping rules

Cross-platform conversion is not a field-name replacement; it must preserve physical semantics. The present implementation uses entity identifiers, geometric coordinates, topological connectivity, linear-elastic material properties, and section properties as primary mapping keys. For nodes and elements, the original identifiers and end-node relationships are retained. For materials and sections, bulk density, elastic modulus, Poisson's ratio, thermal expansion coefficient, outer diameter, and wall thickness are mapped to the corresponding target fields. For result review, Midas DX, DY, and DZ are aligned with SAP2000 U1, U2, and U3, while Midas FX, FY, and FZ are aligned with SAP2000 F1, F2, and F3. The present converter implements automatic mapping only for linear-elastic material parameters and conventional section geometry. It does not establish an equivalent mapping or writeback path for nonlinear material behavior, including fracture energy, bond-slip relations, damage evolution, coupled plasticity, fiber-section constitutive response, or user-defined stress-strain curves. The S2K-to-MGT route can read SAP2000 nonlinear-material tables when building an auxiliary material library, but it records only table presence and selected source fields; the constitutive laws are not converted into MGT material commands. The MGT-to-S2K route has no corresponding nonlinear-material mapping. Therefore, nonlinear material conversion is outside the scope of the current system and must be completed manually in the target software.

In the rule organization, the system uses a combination of static mapping dictionary and dynamic fault-tolerant strategy: conventional fields are directly converted through the rule table, and non-standard fields first enter the cleaning and normalization process; when encountering complex attributes that cannot be reliably mapped, the system retains the original record and marks it as a manual review item. This strategy makes the conversion process both automated and avoids forcibly writing uncertain data into the target model.

The complete data-processing and cross-platform conversion route is shown in Figure 1.

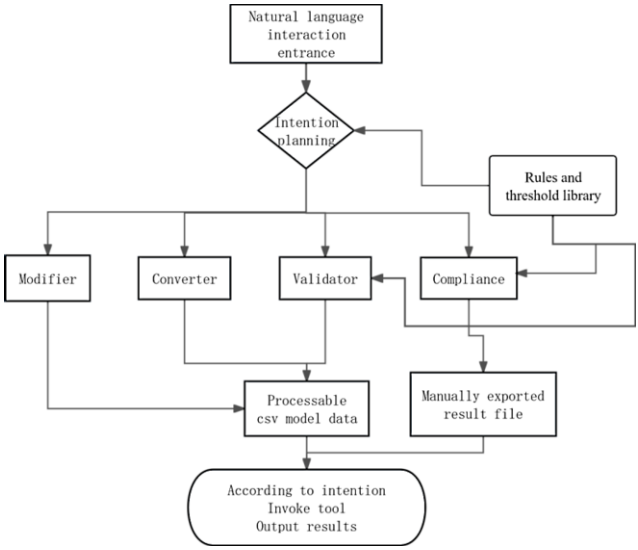


Figure 1. Technology roadmap

2.3. LangGraph agent collaborative workflow

The system uses a LangGraph state machine to convert an engineer's natural-language instruction into an executable tool chain. The workflow contains Planner, Router, Sanitize, and Executor nodes. Planner identifies the task type and target object; Sanitize validates the generated plan, removes unsupported tool calls, and supplements required pre- and post-processing steps; Router distributes the global state to Modifier, Converter, Validator, or Compliance according to the task queue; and Executor invokes deterministic algorithms for file splitting, field modification, format conversion, topology review, and result statistics. In cloud mode, the planner and code-generation components call the model identifier deepseek-chat through an OpenAI-compatible interface with temperature set to 0.1. Task-specific prompts are selected for MGT, S2K, and DXF workflows; they require a JSON plan limited to an explicit tool allowlist and prescribed parameters. The returned plan is sanitized before execution, and a local rule-based planner is used when the cloud request or JSON parsing fails. Thus, the LLM performs intent interpretation and plan generation, whereas numerical modification, file reconstruction, topology search, and engineering checks remain deterministic and reproducible.



Figure 2. LangGraph agent collaboration framework diagram

In the state flow, the system writes the current file path, structured data table, task queue, execution log, exception record and output result into a unified state object. After

each node is completed, the state is updated and returned to the router. The router decides to enter the next tool node, trigger the rollback, or end the process. For the structural model processing task, the key value of this state-based design is that the large language model is only responsible for intent analysis and task planning, and the parts involving numerical modification, file reorganization, topology search and result checking are all performed by deterministic algorithms, thereby reducing the impact of probability generation on engineering results.

2.4. Modifier, Converter, Validator and Compliance module design

The Modifier module is oriented to local model modification. The typical process is split- > code- > merge: the system first splits the .mgt or. s2k file into structured data tables, then generates controlled field operations according to user instructions, and finally reorganizes the modified data table into the target model. In the engineering example, the system can locate the NODE data block and complete the local parameter modification according to the natural language instructions such as 'increasing the Z coordinate of a node by 500 mm'.

Converter module for SAP2000 and Midas cross-platform conversion. The core process is 'split-mapping-recombination-solution-contrast': the source model is split into nodes, elements, materials, sections, constraints and load tables, and then the underlying text of the target software is generated according to the mapping rules and imported into the target software. After the transformation is completed, the system calls Validator and Compliance to review the geometric topology and mechanical response.

The converter's split-mapping-recombination workflow is shown in Figure 3.

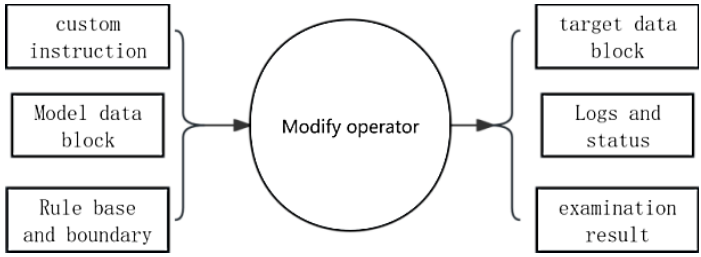


Figure 3. Converter cross-platform conversion logic diagram

The Validator module is oriented towards topology and version conformance reviews. For the single model review, the system identifies the coincidence nodes, repeated elements, isolated nodes, single-ended free rods and double-ended free rods through the node reference relationship and the constraint node set. For the double model review, the system performs version difference according to the primary key of the same number entity, and outputs the results of new addition, deletion, coordinate change, connection relationship change, cross-section change and material change.

Compliance module is oriented to the review of calculation results. The system reads the displacement, stress and internal force result table, extracts the maximum value according to the user's given limit or engineering threshold, counts the overrun object and outputs the distribution result. This module converts the conversion result review from ' looking at the cloud image to find the anomaly ' into a traceable data screening process.

3. Engineering examples and verification schemes

3.1. Validation model and data source

The verification objects include three types of structural models. The first category is the two-way conversion example of SAP2000 and Midas, which is used to test the consistency of nodes, elements, materials, sections and mechanical responses in cross-platform circulation. The second category is the large-span steel structure roof process model, including 11,988 nodes, 24,989 rods and 210 constrained nodes, which is used to verify the identification of topological anomalies. The third category is the homologous.s2k model at different design stages. The benchmark model contains 20,171 nodes, 40,707 bar elements, 23,030 surface elements, 36 material definitions, 1,554 bar section definitions and 32 surface section definitions. The evolution model consists of 49,457 nodes, 53,252 bar elements, and 42,237 surface elements to verify version difference tracking.

3.2. Reverse conversion verification from SAP2000 to Midas

The reverse conversion uses the SAP2000 model as the source file, and the.s2k data block is parsed and mapped to the nodes, elements, sections, and material records required by Midas. Before and after the conversion, the number of nodes remained 761→ 761, the number of rod elements remained 2,888→ 2,888, and there was no node loss or unit redundancy. Under the condition of self-weight, the maximum vertical displacement calculated by Midas is 19.1497 mm, the corresponding vertical displacement calculated by SAP2000 is 19.8647 mm, the peak difference is 0.7151 mm, and the relative difference is 3.73 %. According to the statistics of nodes with absolute displacement not less than 1 mm, the average relative error of U1/ U2 is about 2.03 %, and the average relative error of U3 is about 4.97 %.

The entity-count and vertical-displacement results for this reverse-conversion case are summarized in Table 2.

Table 2. Comparison of Vertical Displacement Results Between Midas and SAP2000

Conversion Direction	Metric	Midas	SAP2000	Difference or Error
SAP2000→Midas	Number of nodes	761	761	0
SAP2000→Midas	Number of frame elements	2,888	2,888	0
SAP2000→Midas	Maximum vertical displacement	19.1497 mm	19.8647 mm	0.7151 mm (3.73%)
Midas→SAP2000	Number of nodes	92	92	0
Midas→SAP2000	Number of frame elements	241	241	0
Midas→SAP2000	Number of area elements	24	24	0
Midas→SAP2000	Maximum vertical displacement	19.9594 mm	20.7564 mm	0.7970 mm (3.84%)

Note: In the SAP2000→Midas case, the nodes with the maximum displacement are concentrated at 446, 457, 666, and 677; in the Midas→SAP2000 case, the node with the maximum displacement is 55.

3.3. Validation of the forward conversion from Midas to SAP2000

The forward conversion takes the 'Kevist ' reticulated shell spatial structure model as an example, and converts the Midas.mgt model into the SAP2000.s2k model. After conversion, the number of nodes is 92→92, the number of rod elements is 241→241, the number of surface elements is 24→24, the materials are Q355 and Q235, and the outer diameters of the circular tube sections of 50×3, P50×4 and 50×5 are consistent with the wall thickness parameters.

Under 1000 times self-weight, the maximum vertical displacement in both software packages occurred at node 55. The SAP2000 and Midas values were 20.7564 mm and 19.9594 mm, respectively, giving an absolute peak difference of 0.7970 mm and a relative error of 3.84%. The mean absolute displacement difference across 92 matched nodes was 0.1093 mm. The vertical support reactions were 79,116.7 kN in SAP2000 and 79,116.72268 kN in Midas, indicating close global vertical equilibrium. Under the unified stress-envelope protocol, the mean and maximum relative differences of the element-level normal-stress envelopes were 2.100% and 7.829%, respectively.

The corresponding support-reaction and stress-envelope comparison results are listed in Table 3.

Table 3. Comparison of Support Reaction and Stress Results

Metric	Midas	SAP2000	Comparison Result
Total support reaction FX	0 kN	0.0 kN	Consistent
Total support reaction FY	0 kN	0.0 kN	Consistent
Total support reaction FZ	79,116.72268 kN	0.0 kN	Difference of approx. 0.02268 kN
Node with maximum vertical displacement	55	55	Consistent
Maximum vertical displacement	19.9594 mm	20.7564 mm	Relative error 3.84%
Average relative difference of element-level stress envelopes	-	-	2.100%
Maximum relative difference of element-level stress envelopes	-	-	7.829%

Note: Support reactions are obtained from the dead-load (self-weight) solution results of both platforms. For stress comparison, corresponding elements are paired under the same load case. The envelope is defined as the maximum absolute normal stress reconstructed from axial force, biaxial bending moments, and the matched section properties at consistent relative locations along each element. The relative difference is calculated as |sigma_SAP - sigma_Midas| / max(|sigma_SAP|, epsilon) x 100%, where epsilon prevents division by a near-zero reference stress. This protocol avoids comparing different stress types, unmatched local-axis directions, or different section positions.

The corresponding SAP2000 and Midas model configurations are compared visually in Figure 4.

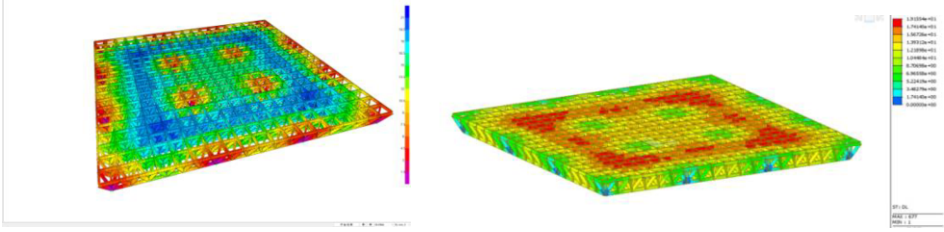


Figure 4. Comparison of the SAP2000 and Midas converted models (mm).

3.4. Topology verification and result review process

The topology verification example adopts the large-span steel structure roof process model. The Validator module first analyzes the node coordinates, member endpoints and constraint node sets, and then identifies single-ended free and double-ended free anomalies based on the number of times the element endpoints are referenced and the constraint state. The Midas result file was used to review the example. The Compliance module reads the displacement, beam element stress and beam element internal force tables, and performs maximum displacement extraction, stress ratio overrun statistics and internal force threshold scanning respectively.

The resulting topology and result-review indicators are summarized in Table 4.

Table 4. Statistical Results of Automatic Review Issues

Review Type	Model Scale or Input	Main Results	Conclusion
Free-end member review	24,989 members, 210 constrained nodes	23 abnormal elements, all are single-end free; 0 double-end free	Topology review failed, targeted correction required
Basic topology check	Same roof process model	0 overlapping nodes, 0 duplicate elements	Good uniqueness of basic numbering and nodes
Displacement check	Displacement result table, limit 0.5 mm	Maximum |DZ| = 18.826 mm, 161 records exceeding displacement limit	Displacement check failed
Stress check	Beam element stress result table, limit 215.0 MPa	5,113 stress over-limit records with Ratio ≥ 1.0, maximum stress ratio 298.0	A large number of over-limit records exist
Internal force check	Beam element internal force result table	2,173 internal force over-limit records, maximum internal force ratio 5.5229×10^7	Manual review of limit values and load case settings required

Note: The limit values and over-limit statistics in the table are derived from the result review output of the verification example.

4. Results and Discussion

4.1. Model transformation integrity analysis

The bidirectional conversion results show that the conversion process based on the underlying text parsing and semantic mapping can maintain the main geometric entities and topological connections. In the example of SAP2000→Midas, the difference between the number of nodes and the number of rod elements before and after the conversion is 0; in the Midas→SAP2000 example, the number of nodes, rod elements and surface elements is consistent, and the corresponding mapping of material and section parameters is also completed. The results show that it is a feasible way to realize the flow of heterogeneous software models by directly analyzing the .s2k and .mgt text

files and reconstructing the structured tables.

Geometric and topological consistency does not guarantee identical local mechanical results. Cross-platform evaluation should therefore use a layered protocol covering entity counts, connectivity, material and section properties, boundary conditions, displacement response, support reactions, member forces, and stress envelopes. In particular, stress comparisons must align the load case, matched element, local coordinate system, section corner, stress type, and position along the element; visual agreement or successful file import alone is not sufficient evidence of mechanical equivalence.

4.2. Comparison of displacement and reaction results

The displacement results show that the maximum vertical displacement positions of the two types of conversion examples are consistent, and the peak relative errors are 3.73 % and 3.84 %, respectively. For the forward conversion example, the average absolute value of the displacement error obtained by point-by-point comparison of 92 coincident nodes is 0.1093 mm, indicating that the converted model can better maintain the overall deformation trend. The comparison of the resultant force of the bearing reaction force further shows that the total vertical reaction force is basically the same in the two software, Midas is 79,116.72268 kN, SAP2000 is 79,116.7 kN.

The stress comparison requires a distinction between global solution consistency and local stress post-processing. For a frame member, normal stress is recovered from axial force and biaxial bending together with section properties. SAP2000 reports P, V2, V3, M2, and M3 at user-defined or automatically added frame output stations, whereas Midas reports axial stress, directional bending stresses, and combined stresses at multiple section corners. Consequently, the maximum stress envelope may be sampled at different longitudinal positions or formed from different corner combinations even when the global equilibrium and member-force fields are close. SAP2000 also reports forces at the faces of supports when end offsets are present, whereas a comparison made at geometric member ends in another platform would not represent the same physical section [17]. Midas defines its combined beam stress as the maximum absolute value among four axial-plus-bending corner combinations and provides interpolated section results for detailed beam analysis [18]. In this study, the reported 7.829% is the largest relative difference after applying the unified normal-stress envelope protocol described in the note to Table 3. It should therefore be interpreted as a localized difference caused primarily by software-specific stress recovery, output-station sampling, and envelope construction, rather than as evidence of a global loss of equilibrium during conversion. This interpretation is supported by the close vertical support reactions and the small mean stress-envelope difference of 2.100%. Nevertheless, if future comparisons show a comparable discrepancy in matched P, M2, and M3 values at the same relative position, the source should instead be traced to model semantics such as section properties, stiffness modifiers, end offsets, releases, or load transfer.

To identify the source of a local stress discrepancy, the comparison should proceed from global to local quantities: first verify reactions and nodal displacements; then compare P, V2/V3, and M2/M3 at the same normalized position x/L; and finally reconstruct stress from the same section properties and the same four physical corner points. Agreement at the force level but disagreement only in the reported stress envelope indicates a post-processing difference, whereas disagreement already present in the forces indicates a model-semantic mismatch.

4.3. Topological anomaly identification and automatic review results

The Validator module identifies 23 single-ended free abnormal members in the roof process model containing 24,989 members, and determines the number of double-ended free abnormalities to be 0. The results have clear engineering significance : single-ended freedom usually corresponds to missing bearings, ineffective connections at the ends of local members or temporary structural residues, which may lead to local geometric variability or unexpected deformation ; double-end freedom is a more serious risk of chain scission, and no such problem is found in this example. Compared with manual item-by-item inspection, batch inspection based on node reference relationship and constrained node set can transform topology risk into a localizable unit list.

The Compliance module converts the result review from the graphical interface cloud view to the limit judgment driven by the data table. The system extracts the maximum vertical displacement of 18.826 mm from the displacement results, and counts 161 displacement overrun records. In the stress results, 5,113 transfinite records with Ratio ≥ 1.0 were detected, and the maximum stress ratio was 298.0. 2,173 transfinite records were detected in the internal force results. The results show that the automatic review module can quickly locate extreme values and overrun objects from the discrete result table, and provide data entry for engineers to review working conditions, limits, component numbers and reinforcement strategies

The reviewed vertical-displacement field is visualized in Figure 5.

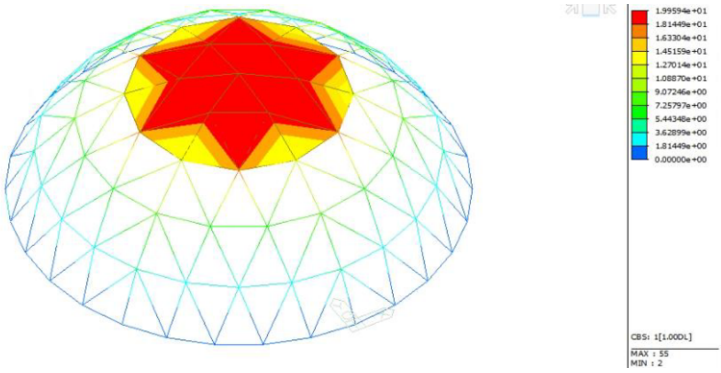


Figure 5. Vertical displacement contour (mm).

4.4. System efficiency and engineering applicability discussion

The engineering test shows that, when the program and large model are accessible, the reconstruction and review process can be completed in approximately 1-2 min. The bidirectional closed-loop conversion of the large model required about 44 s, and the total analysis and transformation time for models containing tens of thousands of nodes and elements was controlled within 360 s. Compared with manual reconstruction, item-by-item checking, and contour-based review, the main contribution of the system is not to replace engineering judgment, but to convert repetitive and error-prone operations into a traceable data workflow.

The current quantitative validation is limited to static self-weight cases and does not establish equivalence for dynamic, lateral-load, staged-construction, or nonlinear analyses. The workflow itself is not restricted to a particular load type because the

planner routes tasks to deterministic data and audit tools; however, future validation should explicitly cover wind and seismic load combinations, response-spectrum and time-history analysis, modal periods and participating mass, base shear, interstory drift, and dynamic force envelopes. Such tests are required before extending the present numerical conclusions beyond the static linear-elastic scope.

At the implementation level, construction-stage nonlinear analysis is also not transferred. Although the code contains a parser for SAP2000 nonlinear-stage definition and stage-data tables, the active conversion entry point does not invoke that parser and removes stage CSV blocks before MGT merging. Accordingly, nonlinear material behavior and nonlinear staged-construction analysis are explicit limitations of the current system, not completed conversion capabilities.

From the perspective of engineering application, the value of LangGraph agent framework is reflected in three aspects. First, users can trigger modification, conversion, verification or review tasks through natural language, reducing the threshold for editing the underlying text directly. Secondly, the deterministic tool chain is responsible for real data analysis and numerical processing, so that the system output can be rechecked. Third, the state machine saves intermediate results, exception logs, and output reports, making it easy for engineers to track the model 's processing path from input to review conclusions.

5. Conclusion

(1) By analyzing the underlying text files of SAP2000.s2k and Midas.mgt, and converting nodes, elements, materials, cross-sections, constraints and result data into structured two-dimensional tables, the geometric topology and physical properties required for cross-platform conversion of structural calculation models can be realized.
(2) LangGraph-based agent collaboration framework can combine natural language intent recognition with deterministic engineering tool chain. The modules of Modifier, Converter, Validator and Compliance are responsible for model modification, format conversion, topology verification and result review respectively, forming a traceable closed-loop workflow.
(3) The engineering examples verify the feasibility of the method. The bidirectional SAP2000-Midas conversion preserved the numbers of nodes and elements. In the forward-conversion case, the maximum vertical-displacement error was 3.84%, the mean relative difference of the element-level stress envelope was 2.100%, and the maximum relative difference was 7.829% under the unified comparison protocol. Topology validation identified 23 single-ended free members. The audit module extracted a maximum vertical displacement of 18.826 mm and detected 5,113 stress-overlimit records.
(4) This method can compress part of the model reconstruction and review process to the minute level, and improve the operability of model difference tracking and overrun object positioning. The follow-up research should be extended to more solvers, nonlinear analysis objects and specification clause mappings, supplemented by benchmarking with manual review and traditional scripts.

References

[1] ZHANG Xiaoyang, HU Zhenzhong. Research on model conversion approach towards structural finite element analysis. Engineering Mechanics, 2017, 34(06): 120-127.
[2] SHEN Lei, LI Jingxue, TAO Yong, et al. Application of Midas Gen to SAP2000 model conversion technology in railway station engineering. Railway Standard Design, 2014, 58(07): 120-124, 130.
[3] YANG Danghui, SU Yuan, SUN Ming. Analysis of data exchange problems in structural design using Revit-based BIM technology. Journal of Information Technology in Civil Engineering and Architecture, 2014, 6(03): 13-18.
[4] DING Jiemin, CHAO Si, FAN Zhong. Review and prospect of high-rise buildings and large-span spatial structures in China. Building Structure, 2019, 49(19): 1-15.
[5] ZHANG Yueqiang, JIAO Chunjie, DING Jiemin. Critical issues and software development research on model conversion from SAP2000 to ABAQUS for complex high-rise structures. Building Structure, 2013, 43(06): 54-57.
[6] BU Weifei, HU Jichao. Brief analysis of the application of SAP and MIDAS software in structural design. Journal of Changchun Institute of Technology (Natural Sciences Edition), 2011, 12(04): 10-12.
[7] LU Xinzheng, LIAO Wenjie, GU Donglian, et al. Research progress on building structural design methods: From simulation-based to artificial intelligence-based. Engineering Mechanics, 2023, 42(3): 1-17.
[8] QIN Sizhong, ZHENG Zhe, GU Yi, et al. Exploring and discussion on the application of large language models in construction engineering. Industrial Construction, 2023, 53(9): 162.
[9] DU S, HOU L, ZHANG G, et al. BIM and IFC data readiness for AI integration in the construction industry: A review approach. Buildings, 2024, 14(10): 3305.
[10] LangChain AI. LangGraph documentation: StateGraph and graph-based agent workflows. https://langchain-ai.github.io/langgraph/
[11] WU Q, BANSAL G, ZHANG J, et al. AutoGen: Enabling next-gen LLM applications via multi-agent conversation. 2023.
[12] WANG L, MA C, FENG X, et al. A survey on large language model based autonomous agents. Frontiers of Computer Science, 2024, 18.
[13] SHINN N, CASSANO F, LABASH B, et al. Reflexion: Language agents with verbal reinforcement learning. Advances in Neural Information Processing Systems, 2023
[14] BESTA M, BLACH N, KUBICEK A, et al. Graph of thoughts: Solving elaborate problems with large language models. In: Proceedings of the AAAI Conference on Artificial Intelligence. 2024.
[15] JIANG J, ZHOU K, DONG Z, et al. StructGPT: A general framework for large language model to reason over structured data. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. 2023.
[16] HUANG J, CHEN X, MISHRA S, et al. Large language models cannot self-correct reasoning yet. arXiv:2310.01798, 2023.
[17] CSI America. SAP2000 Help: Frame Output Stations; Frame Element Internal Forces Output Conventions. https://docs.csiamerica.com/help-files/sap/.
[18] MIDAS Information Technology. MIDAS GEN Online Manual: Force and Stress; Beam Detail Analysis. https://manual.midasuser.com/EN_Common/Gen/.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)

© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261152

Research on the Unsteady Aerodynamic Model of a Conical Probe Using Forced and Free Vibration Based on Parametric Dynamic Derivatives

Hao HAN[a,1,] and Xu LIU[a]
[a] *Naval Submarine Academy, Qingdao, China*
ORCiD ID: Hao HAN, https://orcid.org/0000-0002-7685-2758

Abstract. This paper takes the high supersonic pointed cone standard model as the experimental object, and calculates the unsteady aerodynamic forces under both free vibration and forced vibration conditions. It analyzes the influence of the initial attack angle and rotational inertia on the simulation accuracy, and verifies the applicability of the model to the large attack angle conditions. The results show that under different initial attack angles, the motion time course and hysteresis curve maintain a consistent pattern, exhibiting typical point attractor characteristics; the increase in rotational inertia slows down the amplitude decay, prolongs the vibration period, and increases the number of hysteresis curve loops around the origin; the initial attack angle and rotational inertia have no significant impact on the simulation accuracy.

Keywords. Unsteady aerodynamics; Parametric dynamic derivatives; Hypersonic sharp cone; Free vibration; Nonlinear dynamic characteristics

1. Introduction

As modern aircraft move toward hypersonic speeds and high-maneuverability capabilities, interactions between shock waves and boundary layers, vortex movements, and multi-field coupling render the flow field highly unsteady and nonlinear. These phenomena are likely to trigger adverse outcomes such as sudden changes in aerodynamic loads and attitude instability [1-2]. Precise prediction of unsteady aerodynamic behaviors is essential for overcoming design challenges, making aerodynamic modeling a key research focus—with the modeling of unsteady aerodynamic forces serving as a core area in aerodynamics [3].

However, its substantial computational costs limit its application in engineering contexts such as multidisciplinary collaborative optimization and real-time aeroelastic simulation [4]. Reduced-Order Models (ROMs) replicate core dynamic behaviors via modal extraction and dimensionality reduction, which significantly reduces computational overhead and expands the scope of Computational Fluid Dynamics (CFD) applications [5]. Current mainstream ROMs include Volterra series, Proper Orthogonal

[1] Corresponding author, Hao HAN, Naval Submarine Academy, Qingdao, China. OICID: https://orcid.org/0000-0002-7685-2758.

Decomposition (POD)-Galerkin methods, machine learning surrogates, and neural network models. Although these models can enhance computational efficiency and describe nonlinear flows dominated by static conditions, their accuracy decreases when dealing with strong nonlinear phenomena such as aerodynamic bifurcation and flow field lag [5]. Moreover, their applications are mostly restricted to subsonic/transonic simulations of simple structures (e.g., wings), lacking adaptability to complex hypersonic configurations (e.g., waveriders) and thus failing to meet the design requirements of next-generation aircraft [6].

To address the limitations of ROMs, the CFD-based dynamic derivative solution approach has been adopted for predicting unsteady aerodynamic forces. The classic model proposed by Bryan and Etkin assumes time-invariant dynamic derivatives—a premise that is questionable under strongly unsteady conditions (e.g., hypersonic flight and high angles of attack) [7]. The revised Etkin model incorporates time into baseline parameters, endowing dynamic derivatives with time-varying properties that align with the inherent nature of dynamic stability parameters [8]. In the small angle of attack flight condition, the amplitude of the attitude disturbance of the aircraft around the center of mass is relatively small, and the flow field structure is relatively stable without obvious separation phenomena. At this time, the dynamic derivatives representing the rate of change of aerodynamic forces/moments with respect to motion parameters can be approximated as constants independent of parameters such as angle of attack and Mach number. This simplification assumption greatly reduces the complexity of the aircraft's dynamic modeling, enabling engineers to construct simple state equations through linearization processing, thereby efficiently completing stability analysis, preliminary design of control laws, and other core tasks. In engineering applications of conventional subsonic, transonic, and some supersonic flights, it has extremely high practical value. In sharp contrast, when the aircraft is in a large angle of attack flight condition, the aerodynamic characteristics exhibit significant nonlinear features. On one hand, the relationship between aerodynamic force and angle of attack no longer maintains a simple linear mapping. As the angle of attack increases, the airflow will separate extensively on the aircraft's wing surfaces, rudder surfaces, etc., forming a complex vortex structure, resulting in nonlinear growth or sudden changes in key aerodynamic parameters such as lift, drag, and pitch moment as the angle of attack changes. On the other hand, the effects of time delay and cross-coupling become extremely prominent - the former stems from the lag in the propagation of flow field disturbances and the response of aerodynamic loads, causing the aerodynamic force output to not be synchronized with the attitude changes in real time; the latter is manifested as the coupling and mutual influence between different directions of motion parameters (such as pitch, yaw, roll), for example, the change in roll angular velocity will induce additional yaw moment, further exacerbating the complexity of the dynamic behavior. When the flight speed enters the hypersonic range, the above nonlinear problems become even more challenging due to the intervention of shock-wave - vortex interactions. In the hypersonic flow field, the leading edge of the aircraft will form a strong detached shock wave. The shock wave will interact with the separated vortices on the wing surface, the wake vortices, etc., generating complex interference effects. This interaction not only changes the pressure distribution and energy dissipation laws of the local flow field but also causes violent fluctuations and unsteady oscillations in aerodynamic parameters. Unfortunately, due to the limited depth of current understanding of the non-steady flow mechanism at hypersonic speeds and the accuracy bottleneck of experimental measurement methods in complex flow field environments, the existing numerical simulation models and

engineering empirical formulas are still unable to comprehensively and accurately capture the complex characteristics of the coupled nonlinear aerodynamic forces, time delay, cross-coupling, and shock-wave - vortex interactions under the conditions of large angle of attack and hypersonic flight [9]. This also becomes one of the core bottlenecks restricting the development of dynamic modeling and control technologies for advanced equipment such as high maneuverability aircraft and hypersonic weapons.

Therefore, in response to the aforementioned technical bottlenecks in aerodynamic modeling under conditions of high angle of attack and hypersonic flight, this paper, based on the limitations of existing aerodynamic models, integrates wind tunnel experimental data with high-precision numerical simulation results to systematically optimize the boundary conditions and correction coefficients of the traditional steady-state aerodynamic model; at the same time, it breaks through the simplified assumption that dynamic derivatives are constant at low angles of attack, and innovatively introduces the core concept of parameterized dynamic derivatives, constructing a nonlinear correlation mapping between dynamic derivatives and key flight parameters such as angle of attack, Mach number, and angular velocity. On this basis, this paper fully couples the influence mechanisms of time delay effects and multi-channel cross-coupling terms, relying on the time-domain evolution characteristics of unsteady aerodynamic forces, and develops a set of high-precision unsteady aerodynamic models suitable for large-amplitude attitude motion scenarios. This model, by embedding a dynamic correction module for shock-turbulence interaction, can accurately depict the sudden changes and nonlinear evolution laws of aerodynamic parameters under high-angle-of-attack conditions, effectively filling the accuracy shortcomings of traditional models in complex flight states. [10-11]. Taking a hypersonic sharp cone standard model as the research object, dynamic parameter calculations are performed under free and forced vibration conditions. From three perspectives—vibration type, motion parameters, and computational configuration—the prediction accuracy of aerodynamic forces and flow field parameters between the proposed model and high-precision CFD methods (e.g., DES, LES) is compared to validate the models accuracy and engineering applicability.

2. The mathematical formulation of the model

This paper is based on the core requirements of high-precision numerical simulation of aerodynamic characteristics, and selects the 3D dimensionless N-S equations as the governing equations to describe the fluid motion laws in complex flow fields. This equation eliminates the influence of physical quantities such as characteristic length, characteristic velocity, and characteristic density by dimensionless processing of the continuity equation, momentum conservation equation, and energy conservation equation. It not only significantly improves the convergence and computational efficiency of numerical calculations, but also enhances the universality and comparability of the calculation results of the flow field under different scales and different Mach numbers. At the same time, the three-dimensional form of the equation can completely capture the velocity components, pressure distribution, and energy transfer characteristics of the flow field in the three spatial dimensions, and can accurately reproduce the non-steady flow phenomena such as airflow separation, vortex evolution, and shock-wave-turbulence interaction under high attack angles, providing a solid theoretical and numerical foundation for the subsequent construction and verification of non-steady aerodynamic models. The aerodynamic force/moment

coefficient C_λ is a complex functional that depends not only on the instantaneous values of the angle of attack α and the sideslip angle β, but also on their respective time rates of change, which characterize the transient variation of the aircraft's relative attitude with respect to the incoming flow. Beyond these translational attitude-related parameters, C_λ is further correlated with the angular velocity components along the three orthogonal axes of the body-fixed coordinate system, namely the rolling angular velocity p about the X_b-axis, the pitching angular velocity q about the Y_b-axis, and the yawing angular velocity r about the Z_b-axis. It is crucial to emphasize that this aerodynamic coefficient is not determined by the motion state at a single discrete time point; instead, it is intrinsically associated with the entire temporal evolution history of the aircraft's motion parameters from the initial moment up to the specific time instant t under investigation. Based on the above dependencies, the expression of C_λ can be derived and written in the following form:

$$C_\lambda(t) = C_\lambda(\alpha(\xi), \beta(\xi), p(\xi), q(\xi), r(\xi)) \quad -\infty < \xi \le t \tag{1}$$

To simplify the calculation, only the relationship between C_m and $\alpha(\xi)$ is considered, Eq. (1) can be rewritten as:

$$C_m(t) = C_m(\alpha(\xi)) \quad -\infty < \xi \le t \tag{2}$$

Assume the aircraft is in steady linear flight before $\xi = 0$ with an initial angle of attack α_0. The motion states are defined in two forms:

$$\alpha_1(\xi) = \begin{cases} \alpha(\xi) & 0 \le \xi \le \tau \\ \alpha(\tau) & \xi > \tau \end{cases} \quad \alpha_2(\xi) = \begin{cases} \alpha(\xi) & 0 \le \xi \le \tau \\ \alpha(\tau) + \Delta\alpha & \xi > \tau \end{cases} \tag{3, 4}$$

In the framework of unsteady aerodynamic analysis, the variable ξ is defined as the time coordinate that characterizes the continuous evolution of the flow field and the dynamic response of the aerodynamic system over the entire temporal domain of the study. By contrast, the parameter τ denotes a fixed time that is pre-specified within the time range of the analysis, serving as a reference node to delineate the initial state, a critical transition moment, or a benchmark time point for comparing the unsteady aerodynamic characteristics at different stages of the motion process. When $\xi \le \tau$, $\alpha_1 = \alpha_2$; when $\xi > \tau$, $\alpha_1 - \alpha_2 = \Delta\alpha$. Setting $t = \tau$, the moment coefficients corresponding to the states in Equations (3-4) at any time $\xi \le t$ are $C_{m1}(t) = C_m(\alpha_1(\xi))$ and $C_{m2}(t) = C_m(\alpha_2(\xi))$ according to Eq. (2). When $t > \tau$, $C_{m1}(t) \ne C_{m2}(t)$. Defining $\Delta C_m(t) = C_{m2}(t) - C_{m1}(t)$, if the limit $\lim\limits_{\substack{\Delta\alpha\to 0 \\ t>\tau}} \frac{\Delta C_m(t)}{\Delta\alpha}$ exists uniquely, this limit is defined as the indicial response:

$$A = \lim_{\substack{\Delta\alpha\to 0 \\ t>\tau}} \frac{\Delta C_m(t)}{\Delta\alpha} \tag{5}$$

From Eq. (5), it can be inferred that the aerodynamic moment coefficient difference $\Delta C_m(t)$ under the two motion states is related to $\Delta\alpha$ and the angle α before τ:

$$\Delta C_m(t) = \Delta C_m[\alpha(\xi); \Delta\alpha, \tau, t] \quad -\infty \le \xi \le \tau \tag{6}$$

Based on Eq. (6):

$$A = A[\alpha(\xi); t, \tau] \quad -\infty \le \xi \le \tau \tag{7}$$

According to the definition of the indicial response in Eqs. (6-7), when $t > 0$, the $C_m(t)$ can be written in the following form:

$$C_m(t) = C_m(0) + \int_0^t A\,[\alpha(\xi); t, \tau]\frac{d\alpha}{d\tau}d\tau \quad (0 \le \xi \le \tau) \tag{8}$$

The above equation employs the initial condition: when $\tau < 0$, $\frac{d\alpha}{d\tau} = 0$. Eq. (8) describes the response situation, which not only represents the aircraft's response to the maneuvering actions through control, but also represents the transient process of deviation from the reference flight state due to interference.

To simplify calculations, assume $\alpha(\xi)$ can be expanded into a convergent Taylor series near $\xi = \tau$. Eq. (8) can then be equivalently rewritten as:

$$A[\alpha(\xi); t, \tau] = A\left(t - \tau; \alpha(\tau), \dot{\alpha}(\tau), \ddot{\alpha}(\tau), \cdots\right) \tag{9}$$

In Eq. (9), $t - \tau$ replaces t and τ independently, with a clear physical meaning: given $\tau, \alpha(\tau), \dot{\alpha}(\tau), \ddot{\alpha}(\tau), \cdots$, the change history of $\alpha(\xi)$ is determined, and the transient response depends on $t - \tau$ rather than t and τ separately. Thus, Eq. (8) can be rewritten as:

$$C_m(t) = C_m(0) + \int_0^t A\left(t - \tau; \alpha(\tau), \dot{\alpha}(\tau), \ddot{\alpha}(\tau), \cdots\right)\frac{d\alpha}{d\tau}d\tau \tag{10}$$

By definition of the indicial response, when $t - \tau \to \infty$, $A\left(t - \tau; \alpha(\tau), \dot{\alpha}(\tau), \ddot{\alpha}(\tau), \cdots\right)$ tends to a constant dependent on $\alpha(\tau)$, with the limit $A(\infty; \alpha(\tau))$. The deficiency function is defined as:

$$F\left(t - \tau; \alpha(\tau), \dot{\alpha}(\tau), \ddot{\alpha}(\tau), \cdots\right) = A(\infty; \alpha(\tau)) - A\left(t - \tau; \alpha(\tau), \dot{\alpha}(\tau), \ddot{\alpha}(\tau), \cdots\right) \tag{11}$$

When $t - \tau \to \infty$, $F \to 0$. Substituting Eq. (11) into Eq. (10):

$$C_m(t) = C_m(\infty; \alpha(t)) - \int_0^t F\left(t - \tau; \alpha(\tau), \dot{\alpha}(\tau), \ddot{\alpha}(\tau), \cdots\right)\frac{d\alpha}{d\tau}d\tau \tag{12}$$

With the substitution $u = t - \tau$, Eq. (12) becomes:

$$C_m(t) = C_m(\infty; \alpha(t)) - \int_0^t F\left(u; \alpha(t-u), \dot{\alpha}(t-u), \ddot{\alpha}(t-u), \cdots\right)\frac{d\alpha(t-u)}{du}du \tag{13}$$

Assuming $\alpha(\xi)$ can be expanded into a convergent Taylor series at any point in $[0, t]$ (since τ can be any point in $[0, t]$), $\alpha(t-u), \dot{\alpha}(t-u), \ddot{\alpha}(t-u), \cdots$ (where $t - u = \tau \in [0, t]$) can be expanded into convergent Taylor series near t:

$$\alpha(t-u)=\alpha(t)-\dot{\alpha}(t)u+\frac{\ddot{\alpha}(t)}{2}u^2-\cdots$$

$$\dot{\alpha}(t-u)=\dot{\alpha}(t)-\ddot{\alpha}(t)u+\frac{\dddot{\alpha}(t)}{2}u^2-\cdots$$

$$\ddot{\alpha}(t-u)=\ddot{\alpha}(t)-\dddot{\alpha}(t)u+\frac{\ddot{\alpha}(t)}{2}u^2-\cdots \tag{14}$$

Substituting Eq. (14) into Eq. (13):

$$C_m(t)=C_m(\infty;\alpha(t))+\sum_{i=1}^{\infty}C_i\frac{d^i\alpha(t)}{dt^i} \tag{15}$$

Where:

$$C_i=\frac{(-1)^{i+1}}{(i-1)!}\int_0^t F\left(u;\alpha(t),\dot{\alpha}(t),\ddot{\alpha}(t),\cdots\right)u^{i-1}du \quad (i=1,2,3,\cdots) \tag{16}$$

It can be derived from Eqs. (15-16) that C_i is an explicit function of time t. Based on this, we know that $C_m(t)$ is not only dependent on the generalized state variables but also has an implicit dependence on time t, the specific form of which is shown in Eq. (17).

$$C_m(t)=C_m\left(\alpha(t),\dot{\alpha}(t),\ddot{\alpha}(t),\cdots;t\right) \tag{17}$$

For general motions involving changes in all state variables, the expression for the aerodynamic force/moment coefficient can be similarly derived. ΔC_λ is related to aircraft motion parameters $(\alpha,\dot{\alpha},\beta,\cdots,Ma,h,u,v,\cdots,\delta_e,\dot{\delta}_e)$ and can be expressed as a function of state parameters in Eq. (18):

$$\Delta C_\lambda\left(\alpha,\dot{\alpha},\cdots,\dot{\delta}_e\right)=\frac{\partial C_\lambda}{\partial\alpha}\Delta\alpha+\frac{\partial C_\lambda}{\partial\dot{\alpha}}\Delta\dot{\alpha}+\cdots+\frac{\partial C_\lambda}{\partial\dot{\delta}_e}\Delta\dot{\delta}_e+H.O.T \tag{18}$$

Dynamic derivatives such as $\frac{\partial C_\lambda}{\partial\alpha}$ and $\frac{\partial C_\lambda}{\partial\dot{\delta}_e}$ are obtained through post-processing of time-domain data ΔC_λ. Numerical methods include the integral method, frequency domain transform method, regression method, and phase method.

3. Simulation

This paper conducts wind tunnel experiments based on the sharp cone standard model to investigate the influence of initial angular amplitude and vibration axis on dynamic parameters. Based on the parameter dynamic derivative aerodynamic model, the unsteady aerodynamic forces of free / forced vibration are calculated and compared with

the CFD numerical simulation results to analyze the effect of attack angle and moment of inertia on the simulation accuracy of the model. At the same time, through the forced / free vibration calculations under large attack angle conditions, the performance characteristics of the model within this range are verified.

The calculation model is a 10° sharp cone with a bottom diameter $d = 133.35$ mm (shown in Fig. 1). The computational grid (Fig. 2) is generated using the rigid rotation method, with a first-layer wall-normal spacing of $2 \times 10^{-6}L$. Steady aerodynamic force and dynamic derivative data used in the unsteady aerodynamic model are summarized in Table 1.

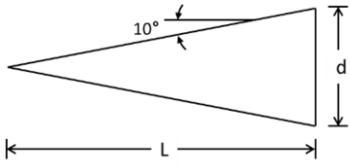


Fig. 1 10 ° sharp cone shape dimension

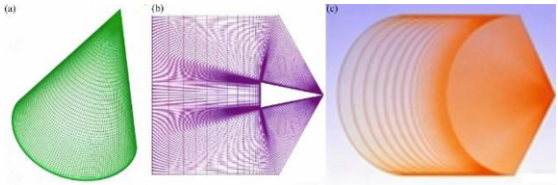


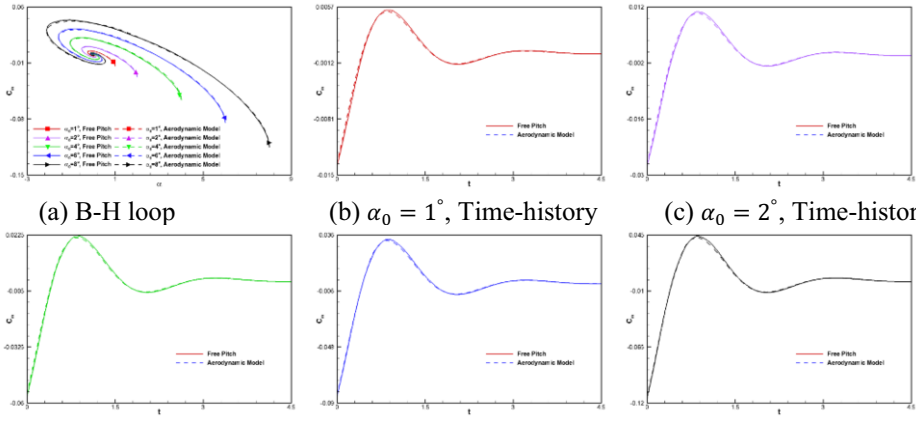
Fig. 2 Pointed cone compute grid; (a) Surface grid (b) Symmetrical surface grid (c) Remote field grid

Table. 1 Comparison of Two Sets of Data of the Pointed Cone

AOA（° ）	C_m	$C_{m\alpha} + C_{mq}$
0	0	-3.396
5	-0.070	-3.197
10	-0.146	-3.236
15	-0.227	-3.986
20	-0.312	-4.415
25	-0.397	-4.469
30	-0.481	-4.539
35	-0.566	-4.618
40	-0.649	-4.431
45	-0.727	-6.110

3.1. Influence of Initial Angle of Attack on Free Vibration

Free pitch vibration simulations are conducted for initial angles of attack ranging from 1° to 8°. The incoming flow Mach number is 4.5, the Reynolds number based on the bottom diameter is 0.36×10^6, and the non-dimensional rotational inertia $I_{zz} = 0.1$. Fig. 3 presents the free vibration hysteresis loops and time histories of the sharp cone for different initial angles of attack. The time histories and hysteresis curves maintain consistent shapes across all initial angles of attack, displaying point attractor characteristics. The aerodynamic model shows good agreement with CFD calculations.



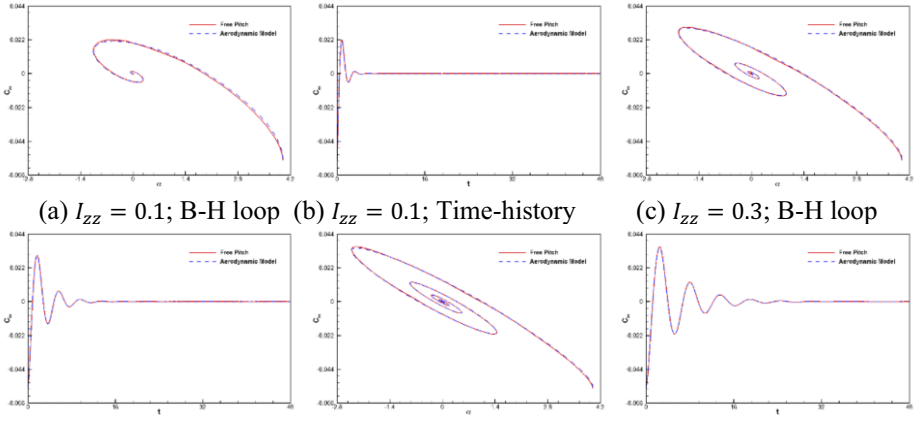
(a) B-H loop (b) $\alpha_0 = 1^\circ$, Time-history (c) $\alpha_0 = 2^\circ$, Time-history

(d) $\alpha_0 = 4^\circ$, Time-history (e) $\alpha_0 = 6^\circ$, Time-history (f) $\alpha_0 = 8^\circ$, Time-history

Fig. 3 Free vibration hysteresis and time history of sharp cone with different initial attack angles

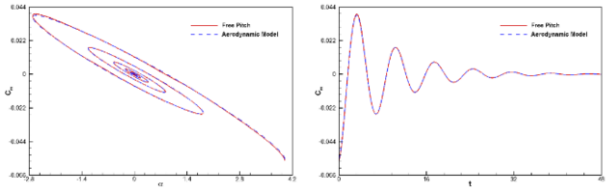
3.2. Influence of Rotational Inertia on Free Vibration

Numerical simulations of free pitch vibration of the sharp cone are performed for rotational inertia I_{zz} values ranging from 0.1 to 1.0. Fig. 4 shows the free vibration hysteresis loops and time histories for different rotational inertias. The frequency of free pitch oscillation is closely related to rotational inertia. From the curve shapes, as I_{zz} decreases, the amplitude decays faster and the vibration period shortens. Hysteresis curves for all rotational inertias exhibit point attractor characteristics. With increasing rotational inertia, the vibration period lengthens and the number of loops around the origin increases. The aerodynamic model is in good agreement with CFD calculations, and rotational inertia has little impact on the simulation accuracy of the aerodynamic model.



(a) $I_{zz} = 0.1$; B-H loop (b) $I_{zz} = 0.1$; Time-history (c) $I_{zz} = 0.3$; B-H loop

(d) $I_{zz} = 0.3$; Time-history (e) $I_{zz} = 0.6$; B-H loop (f) $I_{zz} = 0.6$; Time-history

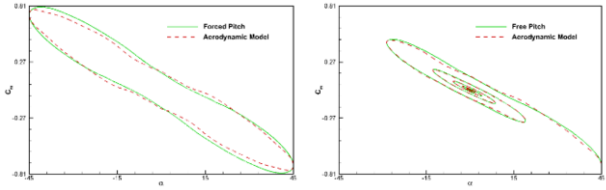


(g) $I_{zz} = 1.0$; B-H loop (h) $I_{zz} = 1.0$; Time-history

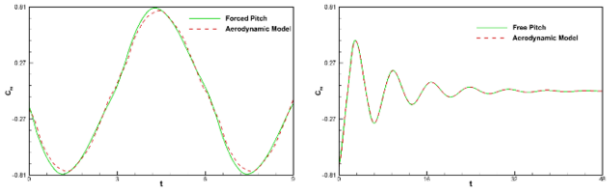
Fig. 4 Free vibration hysteresis and time history of sharp cone with different rotational inertias

3.3. Forced/Free Vibration Calculations Under Large Angle of Attack

Forced/free vibration calculations of the sharp cone under large angle-of-attack conditions are conducted to verify the model's adaptability. The forced pitch vibration mode is $\alpha(t) = \theta = \alpha_0 + \alpha_m \sin(kt)$, with an initial angle of attack $\alpha_0 = 0°$, amplitude of 45°, and reduced frequency $k = 1.0$. For free vibration, the initial angle of attack is 45° and the rotational inertia is 1.0. Fig. 5 shows the large angle-of-attack hysteresis loops and time histories for forced/free vibration. Under high angle-of-attack flight conditions, due to the extensive separation of the airflow on the model surface and the formation of complex vortex structures, there is a significant lag effect between the propagation of flow field disturbances and the response of aerodynamic loads. This effect becomes more pronounced as the angle of attack increases and the vibration frequency rises. At the same time, the time history effect caused by the unsteady evolution characteristics of the flow field will also modulate the amplitude and phase of the unsteady aerodynamic forces by altering the local pressure distribution, energy dissipation patterns, and the temporal characteristics of shock-turbulence interactions. The coupling of these two factors leads to a certain degree of decline in the simulation accuracy of the unsteady aerodynamic force model for high angle-of-attack vibration scenarios compared to those in low angle-of-attack conditions. However, from the overall evolution trend, this model can still accurately capture the changes in dynamic aerodynamic forces with the angle of attack, vibration amplitude, and frequency, effectively reproduce the nonlinear mutations and lag responses of aerodynamic parameters under high angle-of-attack conditions, and provide reliable aerodynamic data support for subsequent flight vehicle dynamics modeling and control law design.



(a) Forced vibration hysteresis loop (b) Free vibration hysteresis loop



(c) Forced vibration time history (d) Free vibration time history

Fig. 5 Large angle-of-attack hysteresis loop and time history of forced/free vibration of sharp cone

4. Conclusion

This paper constructs a non-steady aerodynamic parameter model and conducts forced/freedom vibration calculations for the sharp cone standard model with multiple dynamic parameters. The prediction accuracy of this model and the computational fluid dynamics method in terms of aerodynamic forces and flow field parameters are compared and analyzed. The results show that the established model can effectively calculate the non-steady aerodynamic forces of the sharp cone structure under free and forced vibrations. Its simulation accuracy is not significantly affected by the initial attack angle of free vibration and the rotational inertia I_{zz} of the aircraft body's axis, and it is applicable to forced/freedom vibration conditions with large attack angles. The vibration characteristic analysis reveals that the time history and hysteresis curve morphology of different initial attack angles are consistent, showing a point attractor feature. As I_{zz} increases, the decay rate of vibration amplitude decreases, the period prolongs, and the number of times the hysteresis curve circulates around the origin increases. In summary, the non-steady aerodynamic model based on parameter dynamic derivatives can accurately predict aerodynamic data and flow field parameters under nonlinear conditions.

References

[1] Zhang X., Luo B., et al. Unsteady aerodynamic interference of tilt-quadrotor due to multi-vortex effect // Journal of AerospacePower. 2024. V.40. No.20240216.
[2] Wei X., Wei W., et al. Method of predicting unsteady aerodynamic force based on QPSO-LSTM neural network // Aeronautical Science & Technology. 2025. V.36. No.2. P.102-110.
[3] Wang Z., Wang Y., et al. Construction method of flight data mining model based on LSTM // Aata Aeronautica et Astronautica Sinica. 2021. V.42. No.8. P.525800.
[4] Ronch A., Ghoreyshi M., Badcock K. On the generation of flight dynamics aerodynamic tables by computational fluid dynamics // Progress in Aerospace Sciences. 2011. V.47. P.597-620.
[5] Lucia D., Beran P., Silva W. Reduced-order modeling: new approaches for computational physics // Progress in Aerospace Sciences. 2004. V.40. P.51-117.
[6] Balajewicz M., Nitzsche F., Feszty D. Application of multi-input volterra theory to nonlinear multi-degree-of-freedom aerodynamic systems // AIAA journal. 2010. V.48. P.56-62.
[7] Moni A., Yao W., Malekmohamadi H. Data-driven nonintrusive model-order reduction for aerodynamic design optimization // AIAA Journal. 2024. V.62. No.7 P.1-49.
[8] Zhang W., Zhu J. Unsteady aerodynamic identification of aircraft at high angles of attack // Journal of Tsinghua University. 2017. V.57. P.673-679.
[9] Tang Z., Wang W., et al. Research on the application of aerodynamic models in modern design of aerodynamic experiments // Acta Aerodynamica Sinica. 2017. V.35. P.172-176.
[10] Lei Y., Wenke L., Yuanyuan L. Analysis of boundary condition in prandtl-meyer flow based on riemann invariants // Journal of Wuhan University. 2014. V.60. P.369-373.
[11] Chikitkin A., Petrov M., et al. Parallel versions of implicit LU-SGS method // Lobachevskii Journal of Mathematics. 2018. V.39. P.503-512.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)

© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261153

Multi‑channel Audio Blind Source Separation and Enhancement System Based on Optimized FastICA

Xinyan LI[a], Yan WANG[a], Lanhua ZHANG[a,1]
[a] *College of Medical Information and Artificial Intelligence, Shandong First Medical University and Shandong Academy of Medical Sciences, Tai'an, 271016, China*
ORCiD ID: Lanhua ZHANG https://orcid.org/0000-0003-3895-1170

Abstract. Audio mixing and background noise are pervasive in multi-channel audio acquisition. Traditional FastICA algorithms suffer from imperfect preprocessing, unstable convergence, and severe residual noise after separation, and require manual selection of target components. This paper proposes an integrated blind source separation and enhancement system based on optimized FastICA. Three core modules are designed: a four-step cascaded preprocessing pipeline to optimize convergence stability, an improved spectral subtraction method with iterative feedback to suppress residual noise, and a multi-feature component scoring model for automatic target signal reconstruction. A user-friendly Flask-based web platform is also implemented for plug-and-play processing. Experimental results on 100 test samples show that the proposed system achieves 18.68 dB SNR improvement over the original mixed signal, striking a good balance between separation accuracy and noise suppression, with statistically significant advantages over Audacity, and notable computational efficiency advantages over deep learning-based tools.

Keywords. Independent Component Analysis, FastICA, multi-channel audio separation, spectral subtraction, Web service

1. Introduction

In real-world audio collection, mixed voices and background noise seriously degrade speech quality. Although FastICA [1] is widely used for blind source separation [2], it suffers from imperfect preprocessing, unstable operation and obvious residual noise. Moreover, most existing ICA tools rely on command lines and are not user-friendly.

Aiming at the above limitations, this paper proposes an optimized FastICA-based end-to-end audio separation and enhancement system. The main contributions are: A four-step cascaded preprocessing pipeline is designed to improve FastICA convergence stability. An iterative feedback spectral subtraction method [3] is proposed to suppress separation residual noise and avoid musical noise. A multi-feature component scoring model is built for automatic target component selection, eliminating manual intervention. Besides, a user-friendly Flask-based [4] web platform is developed for

[1] Corresponding Author: Lanhua Zhang, College of Medical Information and Artificial Intelligence, Shandong First Medical University and Shandong Academy of Medical Sciences. E-mail: acm_ict@163.com.

convenient online processing. This paper elaborates the system architecture, algorithm implementation and experimental verification.

2. Method

2.1. Data sources

Three datasets are used for algorithm verification [5]: four 16 kHz single-channel speech clips from the AVSpeech subset, self-generated 4-channel synthetic audio (zero-padding for alignment), and 50 groups of real-world conference recordings for robustness testing.

2.2. Algorithm Models

The system is divided into an audio processing module and a Web service module. The overall workflow is shown in Figure 1.Unlike separate band-pass filtering in prior FastICA audio works, this paper embeds adaptive bandpass into four-step cascaded preprocessing to stabilize convergence without extra post-filter operations.

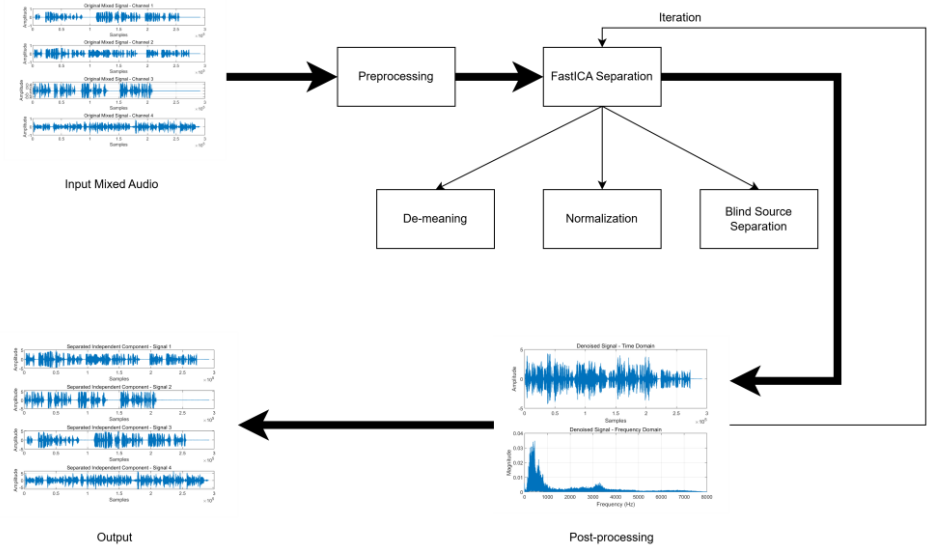


Figure 1. Overall workflow of the proposed system.

A four-step preprocessing strategy is designed to optimize FastICA performance [1]. Let the mixed signal matrix be $X \in \mathbb{R}^{n\times T}$, where n is the number of channels and T is the number of sampling points.

(1) De-meaning: Remove DC component as shown in Eq. (1):

$$X_c = X - \mathbb{E}[\mathbf{X}] \tag{1}$$

(2) Normalization: Balance channel amplitude as shown in Eq. (2), where $\epsilon = 10 - 5$ is a small constant to prevent division by zero.

$$X_{norm} = X_c \oslash (\sigma + \epsilon), \epsilon = 10 - 5 \tag{2}$$

(3) Whitening: Decorrelate signals via covariance matrix decomposition.

(4) Adaptive band-pass filtering: Reserve valid speech band (80‑6000 Hz) and suppress interference.

This work adopts the standard FastICA [1][6] with symmetric orthogonalization. The internal structures of the separation and denoising modules are detailed in Figure 2.The key parameters are set as: maximum iterations = 3000, convergence tolerance = 10^{-10}, activation function, as shown in Eq. (3).

$$g(u) = \tanh(u) \tag{3}$$

An iterative feedback loop is added between separation and post-processing. The denoised signal is re-input for secondary optimization to suppress residual noise.

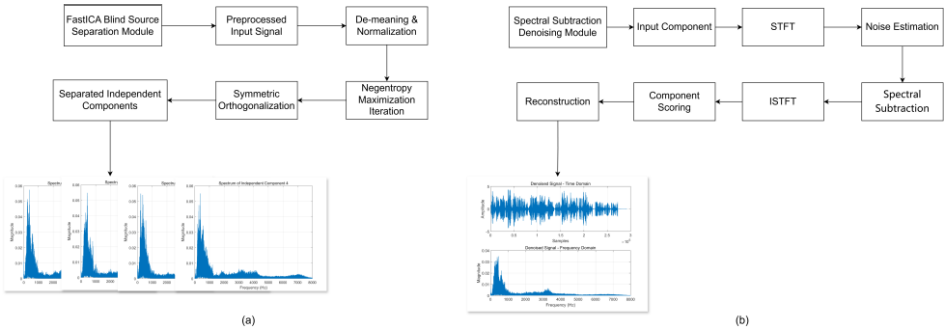


Figure 2. Core algorithm modules: (a) FastICA separation; (b) enhancement and reconstruction.

Each independent component is enhanced separately. First, the signal is transformed via STFT, as defined in Eq. (4):

$$S(f,t) = \sum_{n=-\infty}^{\infty} s(n)\, w(n-t)\, e^{-j2\pi fn} \tag{4}$$

Where $w(n)$ is a Hanning window (length 1024, hop 512). The 20% frames with the lowest energy are used as noise estimate $|N(f)|$, then the improved spectral subtraction is applied [3] as given in Eq. (5):

$$|S_{clean}(f,t)| = \max(|S(f,t)| - \alpha \cdot |N(f)|, \beta \cdot |N(f)|) \tag{5}$$

with α=1.1 (over-subtraction) and β=0.03 (noise floor) to avoid musical noise [7][8]. A component quality scoring model uses three features:

(1) Kurtosis: Reflects the non-Gaussianity of the signal, and speech signals usually have positive kurtosis. Calculated as Eq. (6)

$$kurt_i = \frac{\mathbb{E}[(s_i - \mu_i)^4]}{\sigma_i^4} - 3 \tag{6}$$

(2) Energy: Reflects the intensity of the signal. Defined in Eq. (7)

$$E_i = \sum_t s_i(t)^2 \tag{7}$$

(3) Zero-Crossing Rate: Reflects the smoothness of the signal, and the zero-crossing rate of speech signals is usually in a moderate range. Computed as Eq. (8)

$$zcr_i = \frac{1}{T-1} \sum_{t=1}^{T-1} 1_{\{s_i(t)s_i(t-1)<0\}} \tag{8}$$

The comprehensive scoring formula is given in Eq. (9):

$$\text{score}_i = 0.6 \cdot \text{kurt}_i + 0.35 \cdot E_i - 0.05 \cdot \text{zcr}_i \quad (9)$$

The weights are determined based on speech acoustic properties: kurtosis has the highest weight (0.6) as it directly quantifies non-Gaussianity, the core metric of ICA separation quality; energy (0.35) reflects valid speech intensity; the negative weight for zero-crossing rate (-0.05) suppresses high-frequency noise. This empirical weighting scheme shows stable performance in typical conference speech scenarios.The top-2 components are selected and linearly weighted by normalized scores to reconstruct the enhanced signal, as shown in Eq. (10):

$$s_{\text{enhanced}}(t) = \sum_{i=1}^{2} w_i \cdot s_i(t) \quad (10)$$

Finally, amplitude is normalized to [-0.8,0.8] to prevent clipping. A Flask-based [4][9] Web interface with front-back separation is developed. It supports mainstream audio formats (max 500 MB), real-time progress display, online playback and file download. The core algorithm runs on the MATLAB engine.

3. Result

3.1. Experimental setup

Hardware: Intel Core i7-10700, 16GB RAM, GTX 1660 Ti. Software: MATLAB R2023a, Python 3.9, Flask 2.3.3[4]. The test set contains 100 groups of 4-channel audio (50 synthetic + 50 real-world). SNR and 5-person subjective scoring (1-5, evaluated by 5 listeners as a preliminary assessment) are used for evaluation [10]. Four baselines are selected for comparison: traditional FastICA, Audacity, and representative deep learning separation methods Spleeter [11] and Demucs [12]. Paired t-test and Wilcoxon signed-rank test are adopted for statistical significance verification, with a significance level of α=0.05. SNR was computed after permutation matching and least-squares amplitude scaling for estimated sources to resolve ICA's ambiguities; the mixed signal SNR was computed without scaling.

3.2. Experimental Results

Table 1 lists the average SNR results. The SNR improvement of the proposed system over the original mixed signal and over Audacity, as well as the subjective score differences, all pass significance tests with $p < 0.001$, confirming the improvement is not dataset-specific random variation.

Table 1. Performance Comparison of Different Methods

Method	SNR on synthetic dataset (dB)	SNR on real scene dataset(dB)	Average improvement(dB)	Runtime (seconds/10min audio)
Original mixed signal	-2.69	-2.56	-	-
Traditional FastICA	40.58	0.32	43.27	5
Audacity	3.84	-0.45	6.52	2.8
Spleeter*	36.87	/	/	496.1
Demucs*	29.07	/	/	524.5

Method	SNR on synthetic dataset (dB)	SNR on real scene dataset(dB)	Average improvement(dB)	Runtime (seconds/10min audio)
Proposed system	16.00	2.79	18.68	23.3

*Note: Spleeter [11] and Demucs [12] are included only as reference because they target single-channel music separation.

The proposed system achieves 18.68 dB average SNR improvement over the original mixed signal. Conventional FastICA shows higher objective SNR (40.58 dB) after amplitude scaling, but leaves noticeable residual noise in silent segments (Figs. 3-4). Our system trades moderate numerical SNR for effective noise suppression and full automation, as confirmed by the subjective scores (4.2 vs. 2.8).

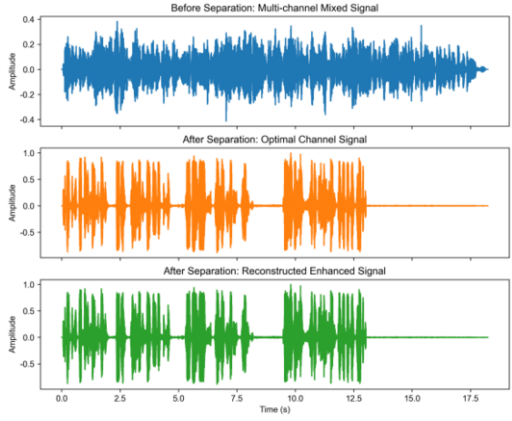


Figure 3. Waveforms before (top), after separation (middle), and after enhancement (bottom). Silent-segment noise is reduced.

As shown in Fig. 3, the time-domain waveform comparison shows that the noise in the silent segments of the audio is significantly reduced. The subjective scoring results from 5 subjects are as follows: 4.2 points for the proposed system, 2.8 points for the traditional FastICA algorithm, and 2.5 points for Audacity software.

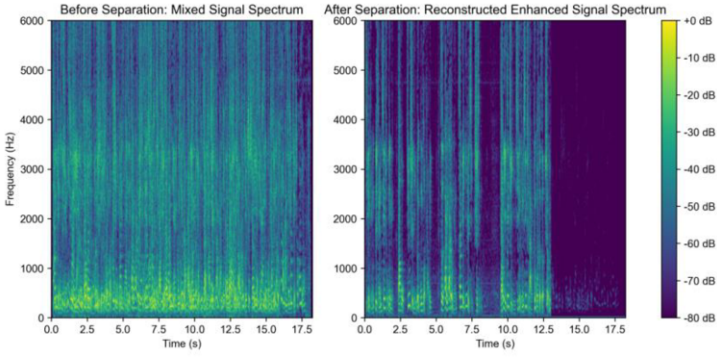


Figure 4. Spectrum comparison before (left) and after (right) separation. Background noise after 12.5 s is removed. (Color bar: energy in dB.)

As shown in Fig. 4, the spectrum comparison results verify that the system can effectively suppress background noise after processing, and the background noise after 12.5 s is almost completely removed.

3.3. Ablation Study

To quantify the contribution of each module, ablation experiments are conducted on 100 synthetic samples. The results are shown in Table 2.

Table 2. Ablation experimental results

Module Configuration	Average SNR (dB)	Improvement over original signal (dB)
Traditional FastICA	40.58	43.27
+ Four-step preprocessing	20.83	23.52
+ Basic spectral subtraction	20.82	23.51
+ Iterative feedback spectral subtraction	18.66	21.35
Complete system (proposed)	16.00	18.68

The SNR decrease after preprocessing is due to the band-pass filter removing out-of-band noise energy that inflated raw FastICA's objective SNR, while preserving the speech band and improving perceptual clarity (confirmed by subjective scores). The four-step preprocessing serves as the core foundation of the system, which significantly optimizes the convergence stability of FastICA. The spectral subtraction modules slightly reduce objective SNR, but effectively suppress residual noise after separation and improve subjective speech clarity. The multi-feature component scoring model realizes automatic target signal selection without accuracy loss, enabling fully automatic end-to-end processing.

4. Discussion

The proposed preprocessing improves FastICA convergence stability [1], while spectral subtraction reduces residual noise and the component scoring model enables automatic target selection. However, the system still has several limitations:

First, it requires input audio to contain at least 2 channels and cannot handle single-channel audio. Second, the algorithm is built on the linear instantaneous mixing assumption. It performs stably in near-field low-reverberation conference scenarios, but separation performance degrades obviously in high-reverberant or long-delay convolutional mixing scenarios. Supplementary tests under T60=0.2/0.5/1.0 s reverberation show average SNR drops by 1.2/3.7/7.5 dB respectively, quantitatively verifying performance degradation under convolutional mixing. Third, the current MATLAB engine limits real-time streaming processing. The offline processing speed is about 26 times faster than real time, which fully meets the needs of batch processing, but low-latency streaming scenarios still need optimization. MATLAB engine cross-call creates major latency; future Python-GPU migration will reduce inference delay for real-time streaming applications. Additionally, the proposed system has longer runtime than raw FastICA (23.3 s vs. 5 s per 10 min audio) due to the additional preprocessing pipeline, iterative feedback denoising, and component scoring steps. This is a reasonable trade-off for fully automatic end-to-end processing and improved output quality.

Future work will investigate single-channel separation, convolutional ICA [13], Python/GPU acceleration, and lightweight deployment for multimicrophone systems.

5. Conclusion

This paper proposes an optimized FastICA-based multi-channel audio blind source separation and enhancement system. It integrates a four-step cascaded preprocessing pipeline, iterative feedback spectral subtraction and a multi-feature component scoring model, realizing fully automatic end-to-end processing. Combined with the Flask-based web platform, it has the advantages of simple operation and strong practicability. Experimental results on 100 test samples show that the system achieves 18.68 dB average SNR improvement over the original mixed signal, striking a good balance between separation accuracy and noise suppression. The system has good application prospects in conference audio processing, speech separation and acoustic monitoring denoising.

Acknowledgements

This research was supported by Shandong Provincial Key R&D Program (Enterprise Innovation Capacity Enhancement Project) grant 2026QYTSGC01005.

References

[1] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," IEEE Transactions on Neural Networks, vol. 10, no. 3, pp. 626-634, 1999.DOI: 10.1109/72.761722

[2] S. Araki, N. Ito, R. Haeb-Umbach, G. Wichern, Z.-Q. Wang, and Y. Mitsufuji, "30+ years of source separation research: Achievements and future challenges," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP), 2025, pp. 1-5.DOI: 10.1109/ICASSP49660.2025.10889006

[3] R. Miyazaki, H. Saruwatari, T. Inoue, K. Shikano, and N. Ono, "Musical-noise-free speech enhancement based on optimized iterative spectral subtraction," in 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2012, pp. 165-168.DOI: 10.1109/ICASSP.2012.6288934

[4] R. Senthilkumar, T. V. Raghavasimhan, R. S. Tejeshwini, C. Kavipriya, and P. Jayanthi, "Real-time suppression of non-stationary noise for web-based calling applications," in *International Conference on ICT for Sustainable Development*, Springer, 2023.

[5] A. Ephrat, I. Mosseri, O. Lang, T. Dekel, K. Wilson, A. Hassidim, and M. Rubinstein, "Looking to listen at the cocktail party: A speaker-independent audio-visual model for speech separation," arXiv preprint arXiv:1804.03619, 2018.DOI: 10.48550/arXiv.1804.03619

[6] Y. Zhao, "Independent component analysis enhancements for source separation in immersive audio environments," M.S. thesis, University of Kentucky, 2013.

[7] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," in Proc. IEEE ICASSP, 1979, pp. 208-211.DOI: 10.1109/ICASSP.1979.1170788

[8] P. C. Loizou, "Reasons why current speech-enhancement algorithms do not improve speech intelligibility and suggested solutions," IEEE Transactions on Audio, Speech, and Language Processing, vol. 19, no. 1, pp. 47-56, 2010.

[9] Q. Nguyen, N. Nguyen, T. Dang, and V. Tran, "Vietnamese Voice2Text: A Web Application for Whisper Implementation in Vietnamese Automatic Speech Recognition Tasks," in Proc. Int. Conf. on Intelligent Systems and Human-Computer Interaction, 2024.

[10] Y. Ephraim, E. Vincent, N. Harlander, and P. Vary, "Subjective and objective quality assessment of audio source separation," IEEE Transactions on Audio, Speech, and Language Processing, vol. 17, no. 5, pp. 950-962, 2009.

[11] R. Hennequin, A. Khlif, F. Voituret, and M. Moussallam, "Spleeter: a fast and efficient music source separation tool with pre-trained models," Journal of Open Source Software, vol. 5, no. 50, p. 2154, 2020.

[12] A. Défossez, N. Usunier, L. Bottou, and F. Bach, "Music source separation in the waveform domain," arXiv preprint arXiv:1911.13254, 2019.DOI: 10.48550/arXiv.1911.13254

[13] A. Brendel, T. Haubner and W. Kellermann, "A Unifying View on Blind Source Separation of Convolutive Mixtures Based on Independent Component Analysis," in IEEE Transactions on Signal Processing, vol. 71, pp. 816-830, 2023, doi: 10.1109/TSP.2023.3255552.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261154

Classification of Sandstone-based Building Materials Based on YOLO Network

Yan Hong [a] and Meng Lv [a,1]
[a] *Nantong Institute of Technology, Nantong, China*
ORCiD ID: Meng Lv https://orcid.org/0009-0003-9916-2498

Abstract. A deep learning classification method based on an improved YOLO network is proposed for the classification of building materials. In the improved YOLO model, the collaboration between downsampling, attention mechanism, and feature extraction network was integrated together, with a focus on key modules such as tagging processing, fast fusion, cross scale connection, and CBAM attention mechanism. During the experiment, CNN method, RNN method, and YOLO method were used as reference methods to classify sandstone building materials. The experimental results fully demonstrate the effectiveness of the improved method, and the classification accuracy of the four sets of experiments is higher than the other three methods.

Keywords. YOLO network, deep learning, building materials, classification

1. Introduction

The classification of building materials in a scientific system is the core foundation for improving the quality and efficiency of the entire process of engineering construction. [1]. The traditional methods, machine learning methods, and deep learning methods for classifying building materials constitute the complete technological evolution path from empirical judgment to intelligent recognition in the current industry [2]. The traditional classification method relies on standardized dimensions such as chemical composition, functional use, and application site, and relies on the professional experience and physical and chemical testing methods of practitioners to complete the classification [3]. By extracting physical and chemical features such as density, strength, and thermal conductivity of building materials, feature vectors are constructed, and training and classification are completed using algorithms such as support vector machines and random forests. It can significantly reduce subjective errors in manual judgment and maintain a high level of recognition accuracy for conventional building materials. However, this method highly relies on manually annotated feature engineering. When facing new building materials with complex textures and high feature overlap, the workload of feature extraction will increase significantly, and the generalization ability will also decrease significantly [4-5]. Deep learning classification methods further break through the limitations of artificial features, relying on models such as convolutional neural networks to automatically mine features from raw data

[1] Corresponding Author: Meng Lv, Nantong Institute of Technology, Nantong, China; Email: lvmeng2026720@126.com.

such as images and spectra of building materials[6]. Without the need for preset feature dimensions, they can accurately capture subtle differences between different building materials, especially suitable for processing complex new building material samples such as photovoltaic integrated panels and phase change energy storage materials [7-8]. On the basis of previous research, we integrated modules such as Label smoothing, weighted fusion, and CBAM, and further integrated an improved bidirectional cross scale connection to better meet the classification of building materials.

2. The Proposed Method

In order to achieve accurate classification of building materials, we adopted an improved deep learning method based on the YOLO framework. The collaborative improvement of downsampling, attention mechanism, and feature extraction network for YOLO series models is the core path to achieve a balance between accuracy and efficiency in complex building material recognition scenarios. In order to reduce overfitting in the YOLO deep learning process, improved labeling processing is applied to each image in the building material image database. Under normal circumstances, there are only two possibilities for whether a building material image belongs to a certain category, as follows:

$$label_i = \begin{cases} 1, i = y \\ 0, otherwise \end{cases} \tag{1}$$

This absolute binary labeling setting can lead to overfitting, which in turn affects the accuracy of the deep learning training process. To solve this problem, we adopted a smoothing coefficient correction processing strategy. The introduction of smoothing coefficients is equivalent to weakening the influence of noise and increasing the consideration of multi classification possibilities. The improved tagging process is as follows:

$$label_i = \begin{cases} 1-\varepsilon, i = y \\ \dfrac{\varepsilon}{k-1}, otherwise \end{cases} \tag{2}$$

Here, k is the total number of categories for all building material images, ε is the smoothing coefficient.

In the process of deep learning, the convolution and pooling results of each channel will ultimately be fused together for classification and discrimination of building material images. There are many commonly used fusion methods, such as Softmax fusion, and so on. Considering the slow process of image processing, a fast fusion method is adopted as follows:

$$O = \sum_i \frac{w_i}{\varepsilon + \sum_j w_j} \times I_i \tag{3}$$

Here, w_i is the weight.

Normalized weighted fusion in deep learning has unique advantages compared to unbounded fusion and Softmax fusion. It abandons the defects of unbounded fusion weight loss and numerical overflow, and also avoids the deployment computing power

loss caused by Softmax fusion index operation. Through simple weight summation normalization, the inference speed can be improved several times without sacrificing feature fusion accuracy. It is more suitable for edge, mobile and other computing power limited scenarios, balancing training stability and hardware deployment friendliness.

Cross scale connection is also an important improvement in our method. During the continuous downsampling process, we further adopted bidirectional feature fusion to complete the connection. Taking the connection from layer four to layer five as an example, the processing procedure is as follows:

$$P_5 = Conv(\frac{w_1 \times F_5 + w_2 \times \mathrm{Re}size(P_6)}{w_1 + w_2 + \varepsilon}) \tag{4}$$

$$N_5 = Conv(\frac{w_1' \times F_5 + w_2 \times F_5 + w_3' \times \mathrm{Re}size(N_4)}{w_1' + w_2 + w_3' + \varepsilon}) \tag{5}$$

Here, N_4 is the fourth stage; N_5 is the fourth stage; F_5 is the input feature map of the fifth layer; P_5 is the prediction header of the fifth layer; P_6 is the prediction header of the sixth layer; w_1, w_2 and w_3 is the input weight parameter.

In order to facilitate a clearer explanation of this cross scale connection process, the following processing structure diagram is provided as Figure 1:

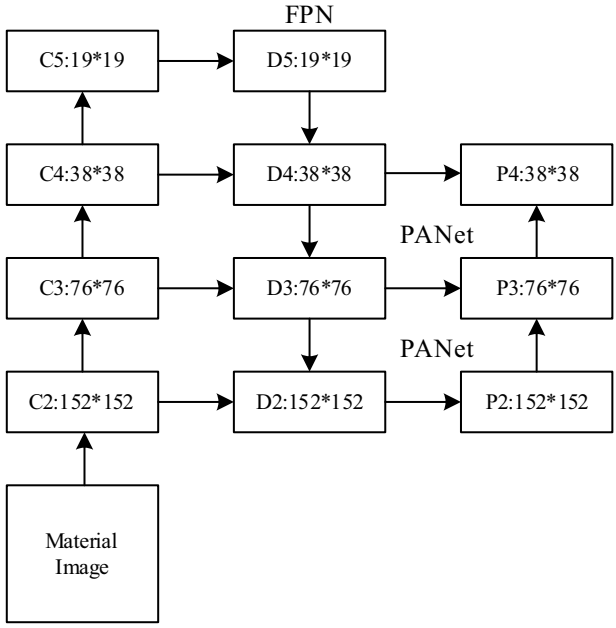


Figure 1. Cross scale connection

To improve the classification performance of deep learning networks on building material images, this paper introduces the mixed attention module CBAM, which synchronously performs dual adaptive calibration on the channel and spatial dimensions of the input model's feature layer, fully exploring the correlation characteristics of features at the channel and spatial levels. The specific structure of this module can be seen in Figure 2.

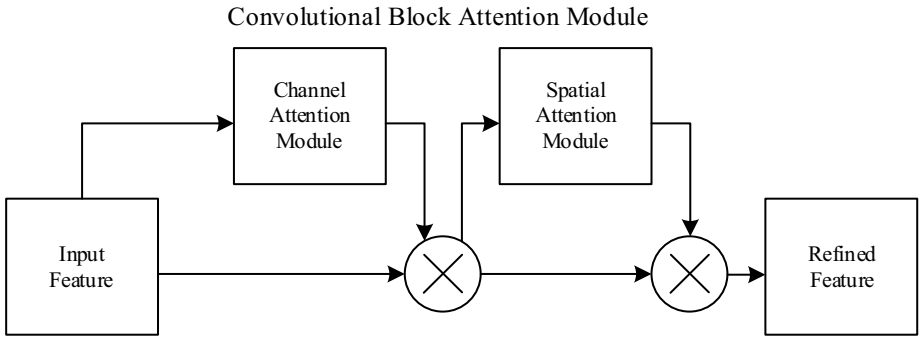


Figure 2. CBAM attention mechanism structure diagram

In the channel dimension processing flow of the CBAM module, the first step is to synchronously perform global average pooling and global maximum pooling operations on the input feature tensor, generating two feature maps with a size of 1 × 1 × c. Then, the two are sent to a multi-layer perceptron fully connected layer with shared weights for feature mapping. In our approach, Label smoothing, weighted fusion, and CBAM are integrated together, and improved bidirectional cross scale connections are incorporated, all of which are improvements to the deep learning processing.

3. Experimental results and analysis

In the experiment, the YOLO11 CLS classification version continues the Backbone Neck Head three-stage architecture, replacing old components with C3k2 and C2PSA modules, and comes with a 640 input size batch 16. The basic configuration with an initial learning rate of 0.01 and 300 iterations can efficiently complete the task of feature extraction and classification of building material images.

During the experiment, sandstone images were used as the research objects and classified based on their differences in particle size. Among them, the particle size diameter of large stones is above 10mm, the particle size diameter of medium stones is between 5mm-10mm, the particle size diameter of small stones is between 1mm-5mm, the particle size diameter of coarse sand is between 0.5mm-1mm, the particle size diameter of medium sand is between 0.35-0.5mm, and the particle size diameter of fine sand is below 0.35mm. Each image contains only one type of sandstone material, and each type of image is collected in approximately equal quantities. The self built sand and stone building material image database has a total of 13000 images, with a ratio of:2:1 between the training set, testing set, and validation set. In the experiment, the validation set was further randomly divided into four groups, with some sample images shown in Figure 3.

The experimental images were captured using the MV-CA030-10GC camera, with a maximum pixel count of 3 million and taken from January to June 2026.The random division of four sets of images may indeed lead to phenomena such as close to frame images in some sets, which also results in differences in the accuracy, recall, and other results of the four experimental groups.

To verify the effectiveness of the proposed method, CNN method, RNN method, and YOLO method were selected as reference methods. Here, CNN adopts the

architecture of "convolution-BN-ReLU-max pooling-fully connected", with parameter configuration: optimizer Adam, initial learning rate of 0.001, Batch Size of 32, Epoch of 10; RNN adopts an architecture of "input 224 dimensions-hidden layer 64 dimensions-fully connected", with parameter: learning rate of 0.001, batch size of 16, and epoch of 20. Four methods were used for the classification of the database shown in Figure 3, and the accuracy of the classification results is shown in Table 1.



Figure 3. Partial images in the database

Table 1. Classification accuracy of four methods (%)

Method	Group1	Group2	Group3	Group4
CNN	90.58	85.71	82.01	84.88
RNN	92.21	86.59	84.08	83.07
YOLO	92.18	87.64	82.11	90.25
OURS	**94.1**	**93.72**	**90.83**	**93.99**

To facilitate a more intuitive presentation of the data in Table 1, it is plotted as a bar chart, as shown in Figure 4.

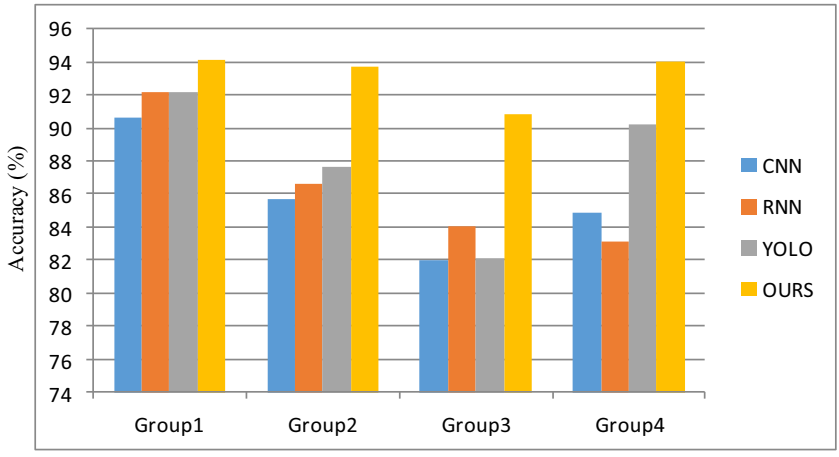


Figure 4. Comparison bar chart of accuracy of four methods

From Figure 4, it can be seen that our method achieved higher accuracy for all four experiments, with significantly better results than CNN, RNN, and YOLO methods, making it more suitable for the classification of building material images.

Further compare the average accuracy, recall, and F1 score of the four methods across the four groups, as shown in Table 2.

Table 2. Comparison of Other Parameters for Four Methods (%)

Method	Precision	Recall	F1
CNN	89.27	91.32	90.08
RNN	87.82	88.76	88.17
YOLO	90.96	95.41	93.67
OURS	**93.32**	**97.66**	**95.12**

The results of ablation study are presented in Table 3.

Table 3. Results of ablation experiment

No	label smoothing	weighted fusion	cross-scale connection	CBAM	Precision (%)
1	√				86.81
2	√	√			88.29
3	√	√	√		91.41
4	√	√	√	√	93.32

4. Conclusions

In this paper, we propose a deep learning classification scheme based on an improved YOLO network. This scheme breaks away from the single module optimization idea of the conventional model, and completes the systematic design for the collaborative adaptation of the down sampling strategy, attention mechanism and feature extraction network. The core processing modules, such as data set refined label labeling, multi-scale feature rapid fusion, shallow and deep characteristics cross layer jump connection, lightweight CBAM attention embedding, are targeted to build. In the comparative experiment, traditional CNN method, time-series feature oriented RNN method, and original YOLO method were selected as three benchmark references. The multi condition sandstone building materials dataset collected on site was used for testing and verification. The classification results under four different interference conditions showed that the improved scheme proposed in this paper had higher classification accuracy than the other three control methods, fully confirming the effectiveness and adaptability of this method in the tested sandstone dataset.

References

[1] Xue S. Building Material Defect Detection and Diagnosis Method Based on Big Data and Deep Learning[J]. Informatica, 2024, 48(16): 1-11.

[2] Sun K , Li Q , Liu Q ,et al.Urban fabric decoded: High-precision building material identification via deep learning and remote sensing[J].Environmental Science and Ecotechnology, 2025, 24(9): 355-364.

[3] Sokolov P E , Sentenberg S A.Application of clusterization algorithms for building materials classification on radioactivity in R[J]. IOP Conference Series Materials Science and Engineering, 2019(1): 687-695.

[4] Yan Q , Jiao F , Peng W .Building Materials Classification Model Based on Text Data Enhancement and Semantic Feature Extraction[J].Buildings, 2024, 14(6):1859-1866..

[5] Banihashemi S , Khalili S , Sheikhkhoshkar M ,et al.Machine learning-integrated 5D BIM informatics: building materials costs data classification and prototype development[J].Innovative Infrastructure Solutions, 2022, 7(3): 822-831.
[6] Bakour F , Chougui A. Beyond Smart Buildings:The Emergence of Intelligent Places through Cutting-Edge Technologies and Material Foundations[J].Journal of Building Material Science, 2025, 7(3):97-117.
[7] Ultralytics. Ultralytics YOLO11 GitHub Repository[EB/OL]. https://github.com/ultralytics/ultralytics, 2024.
[8] Woo S, Park J, Lee J Y, et al. CBAM: Convolutional Block Attention Module[C]//Proceedings of the European Conference on Computer Vision (ECCV), 2018: 1-11.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)

© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261155

A Differential Interbeat Interval-Based Synchronization Method for Long-Term BCG and ECG Signals

Lijuan YANG[a], Haojun HUANG[a], Han ZHANG[a,1] and Baoxian YU[a,2]
[a] *School of Electronic Science and Engineering (School of Microelectronics), South China Normal University, Foshan, China*

Abstract. Ballistocardiogram (BCG) is a promising non-contact cardiovascular monitoring technique, yet its validation must be benchmarked against the gold-standard electrocardiogram (ECG), making BCG-ECG synchronization crucial. In long-term monitoring, asynchronous start times of independent devices and sampling-rate drift of BCG signals can induce initial offsets and cumulative temporal errors. Conventional cross-correlation-based methods cannot directly align raw BCG and ECG waveforms because of their distinct generation mechanisms; therefore, interbeat-interval (IBI) sequences are often used. However, original IBI sequences are quasi-stationary, producing poorly distinguishable correlation peaks and unstable local offset estimation. To address these issues, this paper proposes a BCG-ECG synchronization method that combines cross-correlation of differential interbeat interval (DIBI) sequences with anchor-based resampling. High-quality initial and terminal candidate segments are selected, and J-J and R-R interval sequences and their first-order differences are extracted. Normalized cross-correlation of DIBI features estimates local offsets, and the resulting offsets are used to establish two anchor pairs. The effective BCG sampling rate is then derived from the temporal-span ratio between anchors, followed by resampling and redundant-segment trimming. The method was evaluated on 40 whole-night recordings. Compared with the original IBI baseline, the mean absolute error decreased from 25.91 ± 25.02 ms to 18.15 ± 10.19 ms, and Pearson correlation increased from 0.8261 to 0.9246. These results support the use of DIBI features for improving local offset estimation in long-term BCG-ECG synchronization.

Keywords. BCG; ECG; time synchronization; differential interbeat interval; long-term monitoring

1. Introduction

Ballistocardiogram (BCG) is a non-contact physiological signal reflecting cardiac-induced body vibrations and is suitable for long-term monitoring [1-2]. Electrocardiogram (ECG) serves as a clinical reference for BCG-derived cardiac information. Therefore, accurate BCG-ECG synchronization is required to ensure reliable temporal correspondence between BCG-derived J-J intervals (JJI) and ECG-derived R-R intervals (RRI) [2-3].

[1] Corresponding Author: Han Zhang, zhanghan@scnu.edu.cn
[2] Corresponding Author: Baoxian Yu, yubx@m.scnu.edu.cn

In overnight monitoring, BCG and ECG are often acquired by independent devices, causing asynchronous start and stop times and small sampling-rate mismatches [4]. Asynchrony introduces initial offsets and redundant boundary segments, whereas sampling-rate mismatch can accumulate over long recordings and cause cumulative temporal drift. Thus, synchronization requires both local-offset estimation and temporal-scale correction.

Synchronization approaches include hardware-based, event-based, learning-based, and post-hoc correlation-based methods. For retrospective recordings, post-hoc synchronization is suitable because it requires no shared hardware or predefined common events [5]. Cross-correlation estimates temporal offsets from the lag of maximum correlation between signals or feature sequences.

Because raw BCG and ECG differ markedly in morphology, direct waveform-level correlation is unsuitable for reliable alignment. Their JJI and RRI sequences reflect the same beat-to-beat rhythm and can support local offset estimation [2-3]. Original interbeat-interval (IBI) sequences are quasi-stationary. During slowly varying heart rate, adjacent offsets may yield similar correlations and a broad peak, reducing offset distinguishability. Differential interbeat intervals (DIBI) is therefore used as the matching feature to emphasize beat-to-beat changes and improve correlation-peak distinguishability.

To address these issues, we propose a long-term BCG-ECG synchronization method combining DIBI-based local offset estimation with anchor-based resampling. Two temporally separated anchor pairs are used to estimate the effective BCG sampling rate for temporal-scale correction, followed by redundant-segment trimming. The method was evaluated on 40 whole-night recordings against an original IBI baseline.

2. Proposed Synchronization Method

2.1. DIBI-Based Offset Estimation

DIBI sequences were obtained by taking the first-order differences of the IBI sequences. The original JJI and RRI sequences are usually quasi-stationary, with their values fluctuating around their respective means. When these sequences are directly used for normalized cross-correlation, adjacent candidate offsets may yield similar correlation values, resulting in a poorly distinguishable correlation peak. First-order differencing attenuates the relatively stable component and emphasizes beat-to-beat interval variations. Accordingly, the differential JJI and RRI sequences used for local offset estimation are denoted by JJIV and RRIV, respectively, and defined as

$$\begin{cases} JJIV_n = JJI_{n+1} - JJI_n \\ RRIV_n = RRI_{n+1} - RRI_n \end{cases} \tag{1}$$

the JJI and RRI sequences in Eq. (1) are calculated as

$$\begin{cases} JJI_n = J_{n+1} - J_n \\ RRI_n = R_{n+1} - R_n \end{cases} \tag{2}$$

where J_n and R_n denote the time coordinates of the n-th J peak and R peak, respectively.

The selected JJIV sequence was used as the template and compared with equal-length subsequences within the corresponding RRIV search range to construct the normalized cross-correlation sequence, denoted as *NCC[k]*, where k denotes the shift. The local offset was estimated as the lag corresponding to the maximum normalized cross-correlation value:

$$\hat{\tau} = arg\,max\,NCC\,[k] \tag{3}$$

The estimated offset is subsequently used to determine the corresponding J- and R-peak coordinates for synchronization-anchor establishment.

2.2. Anchor-Based Resampling and Redundant-Segment Trimming

Body movements may interrupt the continuity of usable BCG signals, making continuous JJIV and RRIV sequences difficult to obtain over an entire long-term recording. Therefore, local offset estimation was performed on selected segments. Two temporally separated segments were used to establish two BCG-ECG anchor pairs, each consisting of matched BCG and ECG peak coordinates. Because the coordinate spans between the two BCG anchors and their corresponding ECG anchors represent the same physiological duration, the effective BCG sampling rate was estimated as

$$Fs_{BCG} = \frac{anchor_{BCG_2} - anchor_{BCG_1}}{anchor_{ECG_2} - anchor_{ECG_1}} \cdot Fs_{ECG} \tag{4}$$

Here, $anchor_{BCG_1}$ and $anchor_{BCG_2}$ denote the front and back synchronization-anchor coordinates in BCG, respectively, and $anchor_{ECG_1}$ and $anchor_{ECG_2}$ denote the corresponding coordinates in ECG. Fs_{ECG} denotes the sampling rate of ECG, and Fs_{BCG} denotes the effective BCG sampling rate estimated from the two anchor pairs.

BCG is subsequently resampled according to the estimated effective sampling rate to correct the temporal-scale mismatch. After resampling, redundant segments in BCG and ECG are trimmed according to the synchronization anchors and signal lengths, so that the two signals cover the same physiological time range.

2.3. Synchronization Procedure Using Initial and Terminal Candidate Segments

The initial and terminal portions of each recording were manually screened for two temporally separated candidate segments. Segments with evident motion artifacts, outliers, or persistently poor J-peak detection were excluded, while those with identifiable corresponding JJIV-RRIV variations were preferred.

For each candidate segment, the local offset was estimated using the DIBI-based method in Section 2.1, and equal-length matched RRIV and RRI subsequences were extracted. Candidate quality was assessed using the Pearson correlation coefficient of matched JJIV-RRIV and JJI-RRI pairs, and the cumulative JJI-to-RRI duration ratio. The two correlation thresholds were set to 0.75 and the acceptable time-ratio range to 0.999-1.001. These empirical thresholds were used to exclude segments with weak rhythm consistency or temporal-span mismatch. To reduce dependence on any single metric affected by local peak jitter or signal-quality degradation, a segment was accepted when at least two of the three criteria were satisfied. Otherwise, a new candidate segment was

selected from the corresponding initial or terminal portion. The accepted initial and terminal segments established the two anchor pairs used for the resampling procedure in Section 2.2.

3. Experimental Setup

3.1. Data and Preprocessing

Whole-night sleep recordings were obtained from the Guangzhou Institute of Respiratory Diseases and the Fifth Affiliated Hospital of Sun Yat-sen University. The dataset comprised 40 recordings, including 7 normal, 11 mild, 8 moderate, and 14 severe sleep-apnea cases. BCG was recorded using an under-pillow piezoelectric sensor, while ECG from polysomnography served as the reference. R peaks were detected using the Pan-Tompkins algorithm [6] and manually reviewed when necessary. J peaks were detected using a U-Net-based detector [7], followed by local-maximum refinement and manual correction of evident false-positive and missed peaks. The corrected peak sequences were used to construct RRI, JJI, RRIV, and JJIV for synchronization.

After resampling and trimming, J peaks were redetected from the synchronized BCG and matched to the nearest synchronized R peaks for performance evaluation. These post-synchronization J peaks were not used for offset estimation or anchor establishment. Intervals affected by body movement or out-of-bed activity and IBI values outside 400-2000 ms were excluded from the evaluation.

3.2. Evaluation and Metrics

An original IBI-based synchronization method was used as a controlled baseline to evaluate the contribution of DIBI features. Both methods used the same candidate segments, resampling, trimming, and evaluation procedures, differing only in the feature used for local offset estimation, with DIBI sequences used in the proposed method and original IBI sequences used in the baseline method.

Using ECG-derived RRI as the reference, synchronization performance was evaluated by mean absolute error (MAE), mean relative error (MRE), Pearson correlation coefficient (PCC), concordance correlation coefficient (CCC), and threshold accuracies Pre_{10}, Pre_{30} and Pre_{50}, defined as the proportions of valid paired IBIs with absolute errors below 10, 30, and 50 ms, respectively.

4. Results

Table 1 summarizes the IBI agreement results. Compared with the original IBI baseline, the DIBI-based method reduced MAE from 25.91 $\pm$ 25.02 ms to 18.15 $\pm$ 10.19 ms and increased PCC from 0.8261 $\pm$ 0.2903 to 0.9246 $\pm$ 0.0663. At the record level, the two methods produced identical metrics in 35 of 40 recordings because they selected the same local offsets. In the remaining five recordings, the DIBI-based method improved all seven metrics. After excluding zero differences, a directional exact Wilcoxon signed-rank test yielded $p = 0.0313$ for each metric.

Table 1. Comparison of IBI agreement results using different synchronization features.

Parameter	DIBI-based method	Original IBI method
	Mean ± SD	Mean ± SD
MAE (ms)	18.15 ± 10.19	25.91 ± 25.02
MRE	0.0197 ± 0.0111	0.0285 ± 0.0280
PCC	0.9246 ± 0.0663	0.8261 ± 0.2903
CCC	0.9201 ± 0.0733	0.8209 ± 0.2955
Pre_{10} (%)	57.90 ± 15.39	52.75 ± 21.18
Pre_{30} (%)	86.22 ± 11.14	80.06 ± 21.44
Pre_{50} (%)	90.90 ± 7.98	85.76 ± 17.27

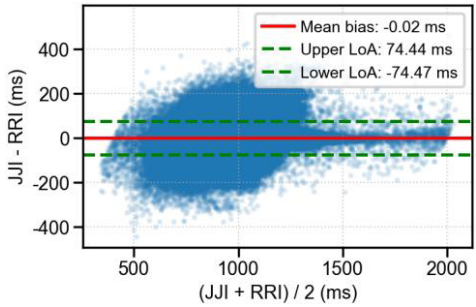


Figure 1. Bland-Altman plot of JJI and RRI after synchronization.

Bland-Altman analysis was performed using the same valid paired IBI set as in Table 1. As shown in Figure 1, 1,055,384 paired IBIs were included after excluding intervals affected by body movement, out-of-bed activity, and abnormal IBI values. The mean JJI-RRI difference was close to zero, indicating no evident systematic bias overall, with 95% limits of agreement from −74.47 to 74.44 ms.

Figure 2 shows a representative recording in which the two features produced different terminal-offset estimates. Both methods yielded an initial offset of 15, whereas the terminal offsets were 11 for DIBI and 43 for original IBI. The original IBI curve showed a broad peak across adjacent offsets, while the DIBI curve showed a more distinct maximum, supporting improved local-offset distinguishability.

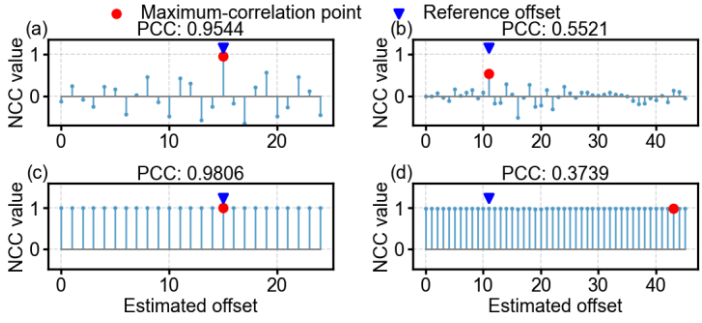


Figure 2. Representative cross-correlation sequences using DIBI and original IBI sequences: (a, b) DIBI sequences in the initial and terminal segments; (c, d) original IBI sequences in the initial and terminal segments.

5. Conclusion

This study proposed a long-term BCG-ECG synchronization method combining DIBI-based local offset estimation with anchor-based resampling. Compared with the original IBI baseline, DIBI improved IBI agreement and the distinguishability of local correlation peaks. Anchor-based resampling was used to correct the BCG temporal scale, followed by redundant-segment trimming to align the valid signal ranges.

The method depends on reliable peak detection and candidate-segment quality. Manual segment selection may limit scalability and introduce operator dependence, while two anchors may be insufficient for substantial nonlinear or time-varying drift. Further validation across larger populations, devices, and signal-quality conditions is needed. Future work will investigate automated candidate selection and multi-anchor segment-wise correction.

Declaration

The authors used ChatGPT (OpenAI) and DeepSeek for language refinement throughout the manuscript, structural organization of selected sections, and preliminary literature organization. These tools were not used to generate research ideas, data, or results. All AI-assisted content was reviewed and approved by the authors, who take full responsibility for the final manuscript.

References

[1] Inan OT, Migeotte PF, Park KS, et al. Ballistocardiography and seismocardiography: a review of recent advances. IEEE J Biomed Health Inform. 2015;19(4):1414-27.
[2] Qiu J, Lyu T, Liu L, et al. A ballistocardiogram dataset with reference ECG signals for bed-based heart rhythm assessment. Sci Data. 2025;12(1):1651.
[3] Yanagisawa N, Yao B, Zhang J, et al. Comparative analysis of heart rate variability indices from ballistocardiogram and electrocardiogram: a study on measurement agreement. Heart Vessels. 2025;40(7):655-64.
[4] Wang H, Li J, McDonald BE, et al. Comparison between two time synchronization and data alignment methods for multi-channel wearable biosensor systems using BLE protocol. Sensors (Basel). 2023;23(5):2465.
[5] Shaabana A, Zheng R. CRONOS: a post-hoc data driven multi-sensor synchronization approach. ACM Trans Sens Netw. 2019;15(3):26:1-26:20.
[6] Pan J, Tompkins WJ. A real-time QRS detection algorithm. IEEE Trans Biomed Eng. 1985;BME-32(3):230-6.
[7] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, et al., editors. Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015. Cham: Springer; 2015. p. 234-41.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)

© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA261156

Aircraft Leasing Strategies in the Context of the Low-Altitude Economy via an Evolutionary Game and Simulation Approach

Wenchao Fang[a,1]
[a] *Guangdong Polytechnic of Industry and Commerce, China*

Abstract. Aircraft leasing is an important component of developing the low-altitude economy. Lessors and lessees are two important participants in aircraft leasing. This paper constructs an evolutionary game model between lessors and lessees, analyzes how the two parties choose strategies in aircraft leasing, and further investigates the impact of parameter changes in the game model on the evolutionary outcomes. The results show that the system can achieve an evolutionarily stable strategy, which corresponds to three different stages of aircraft leasing: the embryonic stage, the growth stage, and the mature stage. Key factors affecting the evolutionary outcomes include the cost of strict supervision by lessors, the cost of active maintenance by lessees, the allocation of liability for aircraft accidents, and the probability that accidents are discovered by the public or exposed by the media. By effectively reducing the cost of strict supervision by lessors, lowering the cost of active maintenance by lessees, clarifying the allocation of liability for aircraft accidents, and strengthening public and media supervision of the leasing market, the high-quality development of the aircraft leasing market can be promoted.

Keywords. Low-altitude economy, leasing, lessor, lessee, evolutionary game

1. Introduction

The low-altitude economy is an emerging economic activity underpinned by the development of low-altitude flight and formed through segments such as aircraft manufacturing, leasing, and operations. Aircraft leasing is an important component of developing the low-altitude economy and is crucial to the quality of its development. Therefore, aircraft leasing has become a focal point of the low-altitude economy.

At present, academia has conducted a series of studies on aircraft leasing. Marintseva and Athousaki [1] investigated the key performance indicators and leasing efficiency of the aircraft leasing industry from the perspective of airlines. Sun *et al.* [2] compared the characteristics and differences of aircraft leasing policies in Shanghai and Tianjin and proposed corresponding suggestions. Wandelt *et al.* [3] adopted a literature review method to analyze the major challenges facing the aircraft leasing industry. Burdina *et al.* [4] studied methods for calculating aircraft lease payments. Lin *et al.* [5] used system dynamics to study the sustainable development of the aircraft leasing

[1] Corresponding Author, Wenchao Fang, Guangdong Polytechnic of Industry and Commerce, China; Email: fangwc163@126.com

industry. Dissanaike *et al.* [6] examined the benefits or costs to suppliers, airline shareholders, and passengers in aircraft leasing. Kuhle *et al.* [7] developed a blockchain-based decentralized digital asset management system for commercial aircraft leasing. Bourjade *et al.* [8] empirically examined the impact of aircraft leasing choices on the financial performance of airlines.

In terms of game-theoretic research on the leasing market, Li *et al.* [9] constructed Stackelberg game and cooperative game models to study peer-to-peer power trading and pricing in community microgrids with leased shared energy storage. Zhu *et al.* [10] constructed a Stackelberg game model and proposed a shared energy storage leasing operation scheme for multiple wind farms. Qiao *et al.* [11] established a Stackelberg game model to study the scheduling of an active distribution network in which a microgrid group leases shared energy storage. Zhang and Huang [12] constructed a Stackelberg game model and analyzed the leasing pricing and capacity allocation of shared energy storage in a distribution network system with multiple microgrids. Li *et al.* [13] constructed a master-slave game model to study the rental problem of optimal allocation of shared energy storage. Li and Wang [14] constructed a Stackelberg game model to study the differences between the long-term lease battery swapping mode and the one-off purchase battery swapping mode for electric vehicles. Zhu *et al.* [15] constructed a Stackelberg game model to study electric vehicle rental fee schemes.

In summary, existing literature mainly focuses on aircraft leasing performance, policies, and benefits. Moreover, most existing studies on the leasing market mainly adopt methods such as Stackelberg games, while studies using evolutionary games to investigate aircraft leasing are relatively rare. Evolutionary games have been widely applied in the transportation field [16−20]. Stackelberg games are analyzed based on complete rationality. Since the two major parties involved in aircraft leasing, namely lessors and lessees, are constrained by subjective and objective conditions and often possess only bounded rationality, an evolutionary game based on bounded rationality is more consistent with the actual situation of the aircraft leasing market. Therefore, this paper attempts to conduct an evolutionary game study of aircraft leasing strategies in the context of the low-altitude economy and, based on numerical examples, carries out simulation analysis to derive corresponding conclusions.

2. Basic Assumptions and Model Construction

In the process of aircraft leasing strategy decision-making, lessors and lessees are the two participants in the evolutionary game, and both are bounded-rational agents. The basic assumptions of this game are as follows.

Assumption 1: The lessor′s strategies are "strict supervision" or "loose supervision". Strict supervision means that the lessor complies with the provisions of the leasing contract and strictly supervises the lessees use of the aircraft, with relatively high supervision costs. Loose supervision means that the lessor does not comply with the provisions of the leasing contract and loosely supervises the lessee′s use of the aircraft, with relatively low supervision costs. The lessee′s strategies are "active maintenance" or "passive maintenance". Active maintenance means that the lessee maintains the aircraft strictly in accordance with the leasing contract and the actual use of the aircraft; the cost of active maintenance is relatively high, but the risk of aircraft accidents is relatively low. Passive maintenance means that the lessee does not maintain the aircraft in accordance

with the leasing contract and the actual use of the aircraft; the cost of passive maintenance is relatively low, but the risk of aircraft accidents is relatively high.

Assumption 2: The proportion of lessors choosing strict supervision is x, and the proportion choosing loose supervision is $1-x$. The proportion of lessees choosing active maintenance is y, and the proportion choosing passive maintenance is $1-y$.

Assumption 3: When the lessor chooses the strict supervision strategy, the additional benefit obtained from implementing strict supervision is R_1, and the strict supervision cost borne is C_1. When the lessor chooses the loose supervision strategy, no additional benefit is obtained, and the loose supervision cost borne is 0. The rental income obtained by the lessor from leasing the aircraft is D.

Assumption 4: When the lessee chooses the active maintenance strategy, the cost borne for active maintenance is C_2. The basic benefit obtained by the lessee from operating business by leasing the aircraft is R_2, and the passive maintenance cost borne is 0.

Assumption 5: When the lessor chooses the strict supervision strategy, if the lessee chooses passive maintenance, the lessor imposes a penalty on the lessee, with a fine of F. When the lessor chooses the loose supervision strategy and the lessee chooses the passive maintenance strategy, because the aircraft use is not effectively supervised and maintained, the loss caused by accidents due to improper aircraft use (including compensation to consumers by the lessor and lessee, etc.) is B, where the lessor′s share of the loss is s, and the lessee′s share is $1-s$; s and $1-s$ can be regarded as the proportions of liability borne by the lessor and lessee in accident compensation. When the probability that such behavior is discovered by the public or exposed by the media is t, the public discovery or media exposure has a negative impact on the corporate image and social influence of both the lessor and the lessee, and the losses of the lessor and lessee are H_1 and H_2, respectively.

Assumption 6: The lessee cannot sublease the leased aircraft to other individuals or organizations.

Assumption 7: All the above parameters are greater than zero.

Based on the above assumptions, the payoff matrix of the evolutionary game between the aircraft lessor and lessee is constructed, as shown in Table 1.

Table 1. The payoff matrix of the evolutionary game between aircraft lessors and lessees.

		Lessee	
		Active maintenance y	Passive maintenance $1-y$
Lessor	Strict supervision x	R_1-C_1+D, R_2-C_2-D	R_1-C_1+F+D, R_2-F-D
	Loose supervision $1-x$	D, R_2-C_2-D	$-tH_1-sB+D$, $R_2-tH_2-(1-s)B-D$

3. Evolutionary Game Analysis of Aircraft Lessors and Lessees

3.1 Equilibrium points of the evolutionary process

Let E_{11} denote the expected payoff of the lessor when choosing the strict supervision strategy, as shown below.

$E_{11}=y(R_1-C_1+D)+(1-y)(R_1-C_1+F+D)$ (1)

Let E_{12} denote the expected payoff of the lessor when choosing the loose supervision strategy, as shown below.

$E_{12}=yD+(1-y)(-tH_1-sB+D)$ (2)

Therefore, the average payoff of the lessor is E_1, as shown below.

$E_1=xE_{11}+(1-x)E_{12}$ (3)

The replicator dynamic equation for the lessor′s strict supervision is shown below.

$$F(x)=\frac{dx}{dt}=x(E_{11}-E_1)=x(1-x)[R_1-C_1+(1-y)(F+tH_1+sB)] \quad (4)$$

Let E_{21} denote the expected payoff of the lessee when choosing the active maintenance strategy, as shown below.

$E_{21}=x(R_2-C_2-D)+(1-x)(R_2-C_2-D)$ (5)

Let E_{22} denote the expected payoff of the lessee when choosing the passive maintenance strategy, as shown below.

$E_{22}=x(R_2-F-D)+(1-x)(R_2-tH_2-(1-s)B-D)$ (6)

Therefore, the average payoff of the lessee is E_2, as shown below.

$E_2=yE_{21}+(1-y)E_{22}$ (7)

The replicator dynamic equation for the lessee′s active maintenance is shown below.

$$F(y)=\frac{dy}{dt}=y(E_{21}-E_2)=y(1-y)[-C_2+xF+(1-x)(tH_2+B-sB)] \quad (8)$$

By setting $F(x) = 0$ and $F(y) = 0$, a two-dimensional dynamical system can be obtained simultaneously. Solving it yields the local equilibrium points of the system: (0,0), (0,1), (1,0), (1,1), and (x, y), where $x^* =[C_2 - tH_2 - (1-s)B]/[F - tH_2 - (1-s)B]$, $y^* = (F + tH_1 + sB + R_1 - C_1)/(F + tH_1 + sB)$.

3.2 Stability analysis of equilibrium points

To determine whether the equilibrium points are stable, a local stability analysis of the system's Jacobian matrix can be performed.

The Jacobian matrix (J) of the system is as shown below.

$$J=\begin{bmatrix}\frac{\partial F(x)}{\partial x} & \frac{\partial F(x)}{\partial y}\\ \frac{\partial F(y)}{\partial x} & \frac{\partial F(y)}{\partial y}\end{bmatrix}=\begin{bmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{bmatrix} \quad (9)$$

$a_{11}=(1-2x)(R_1-C_1+F+tH_1+sB-yF-ytH_1-ysB)$ (10)

$a_{12}=x(1-x)(-F-tH_1-sB)$ (11)

$a_{21}=y(1-y)(F-tH_2-B+sB)$ (12)

$a_{22}=(1-2y)[-C_2+xF+(1-x)(tH_2+B-sB)]$ (13)

If the following conditions are simultaneously satisfied, the equilibrium point of the replicator dynamic equations is stable, and this equilibrium point is the evolutionarily stable strategy (ESS) of the game.

(1) $trJ=a_{11}+a_{22}<0$ (trace condition)

(2) $detJ=a_{11}a_{22}-a_{12}a_{21}>0$ (determinant condition)

According to the specific values of the five local equilibrium points, the trace of (x^*, y^*) is 0, which does not satisfy the conditions for an evolutionarily stable equilibrium. Therefore, the local equilibrium point (x^*, y^*) is not an evolutionarily stable strategy. Furthermore, the specific values of the other local equilibrium points can be obtained, as shown in Table 2.

Table 2. Specific values of the local equilibrium points.

Equilibrium point	a_{11}	a_{12}	a_{21}	a_{22}
(0,0)	$R_1-C_1+F+sB+tH_1$	0	0	$B-C_2-sB+tH_2$
(0,1)	R_1-C_1	0	0	$-B+C_2+sB-tH_2$
(1,0)	$-R_1+C_1-F-sB-tH_1$	0	0	$F-C_2$
(1,1)	$-R_1+C_1$	0	0	$-F+C_2$

Proposition 1: When $R_1-C_1+F+sB+tH_1<0$, $B-C_2-sB+tH_2<0$, $R_1-C_1<0$, and $F-C_2>0$, the ESS of the system is (0,0), and the corresponding strategy combination is (loose supervision, passive maintenance).

Proof: When $R_1-C_1+F+sB+tH_1<0$, $B-C_2-sB+tH_2<0$, $R_1-C_1<0$, and $F-C_2>0$, the evolutionary stability analysis of the system is shown in Table 3.

Proposition 1 indicates that when the cost of strict supervision by the aircraft lessor is greater than the sum of the additional benefit obtained from strict supervision, the penalty received, the loss caused by aircraft accidents, and the loss caused by public discovery or media exposure, and is also greater than the additional benefit obtained from strict supervision—that is, when the cost of strict supervision is sufficiently high—the lessor tends to choose the loose supervision strategy. When the cost of active maintenance by the aircraft lessee is greater than the sum of the loss caused by aircraft accidents and the loss caused by public discovery or media exposure, but less than the penalty it pays, the lessee realizes that the lessor chooses the loose supervision strategy. At this point, since the cost of active maintenance is relatively high, the lessee also tends to choose the passive maintenance strategy.

Table 3. Stability analysis of the equilibrium points in Proposition 1.

Equilibrium point	*trJ*	*detJ*	Stability
(0,0)	–	+	ESS
(0,1)	+	+	Unstable point
(1,0)	Uncertain	–	Saddle point
(1,1)	Uncertain	–	Saddle point

Similarly, the following propositions can be obtained. The proof process is consistent with that of Proposition 1 and is therefore omitted.

Proposition 2: When $B-C_2-sB+tH_2>0$, $R_1-C_1<0$, and $F-C_2>0$, the ESS of the system is (0,1), and the corresponding strategy combination is (loose supervision, active maintenance).

Proposition 2 indicates that when the cost of active maintenance by the aircraft lessee is less than the sum of the loss caused by aircraft accidents and the loss caused by public discovery or media exposure, and is also less than the penalty it pays—that is, when the cost of active maintenance is sufficiently low—the lessee tends to choose the active maintenance strategy. When the cost of strict supervision by the aircraft lessor is greater than the additional benefit obtained from strict supervision, the lessor observes the lessee's active maintenance behavior. At this point, since the cost of strict supervision is relatively high, the lessor will relax supervision and tend to choose the loose supervision strategy.

Proposition 3: When $R_1-C_1+F+sB+tH_1>0$, $R_1-C_1>0$, and $F-C_2<0$, the ESS of the system is (1,0), and the corresponding strategy combination is (strict supervision, passive maintenance).

Proposition 3 indicates that when the cost of strict supervision by the aircraft lessor is less than the sum of the additional benefit obtained from strict supervision, the penalty

received, the loss caused by aircraft accidents, and the loss caused by public discovery or media exposure to the lessor and lessee, and is also less than the additional benefit obtained from strict supervision—that is, when the cost of strict supervision is sufficiently low—the lessor tends to choose the strict supervision strategy. When the cost of active maintenance by the aircraft lessee is greater than the penalty it pays—that is, when the cost of active maintenance is sufficiently high—the lessee tends to choose the passive maintenance strategy.

Proposition 4: When $R_1–C_1+F+sB+tH_1>0$, $R_1–C_1>0$, and $F−C_2>0$, the ESS of the system is (1,1), and the corresponding strategy combination is (strict supervision, active maintenance).

Proposition 4 indicates that when the cost of strict supervision by the aircraft lessor is less than the sum of the additional benefit obtained from strict supervision, the penalty received, the loss caused by aircraft accidents, and the loss caused by public discovery or media exposure to the lessor and lessee, and is also less than the additional benefit obtained from strict supervision—that is, when the cost of strict supervision is sufficiently low—the lessor tends to choose the strict supervision strategy. When the cost of active maintenance by the aircraft lessee is less than the penalty it pays—that is, when the cost of active maintenance is sufficiently low—the lessee tends to choose the active maintenance strategy.

4. Numerical Simulation Analysis of the Evolutionary Game

To more directly observe and analyze the interactive game process between the two parties—aircraft lessors and lessees—this paper employs MATLAB software to conduct a simulation analysis of the evolutionary game.

4.1 Simulation analysis of evolutionarily stable equilibria

With the parameters set as listed in Table 4, the corresponding preconditions of Propositions 1–4 are satisfied. Through numerical simulation, it can be seen from Figures 1–4 that Propositions 1–4 hold. The initial probabilities of the lessor choosing the strict supervision strategy and the lessee choosing the active maintenance strategy are both set to 0.5.

Table 4. Parameter settings for Propositions 1–4.

Proposition	R_1	C_1	C_2	F	s	B	t	H_1	H_2	ESS	Simulation result
Proposition 1	2	9	3	4	0.5	3	0.5	2	2	(0,0)	Figure 1
Proposition 2	2	5	2	4	0.5	3	0.5	2	2	(0,1)	Figure 2
Proposition 3	2	1	2	1	0.5	3	0.5	2	2	(1,0)	Figure 3
Proposition 4	2	1	2	3	0.5	3	0.5	2	2	(1,1)	Figure 4

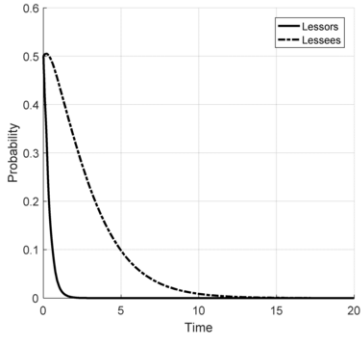


Figure 1. Evolutionary simulation results for Proposition 1.

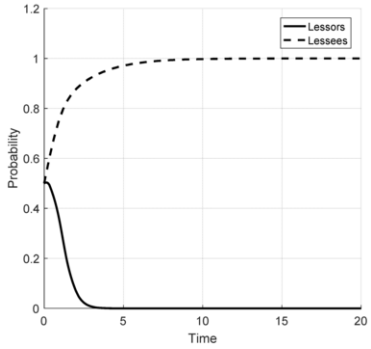


Figure 2. Evolutionary simulation results for Proposition 2.

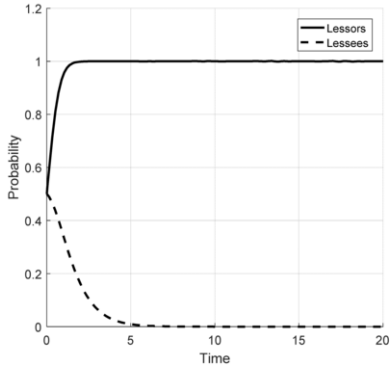


Figure 3. Evolutionary simulation results for Proposition 3.

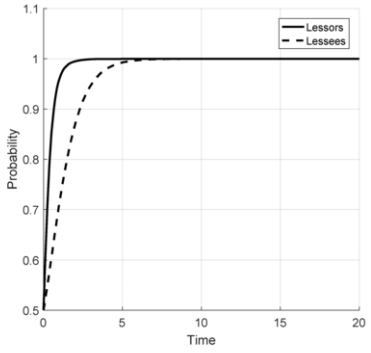


Figure 4. Evolutionary simulation results for Proposition 4.

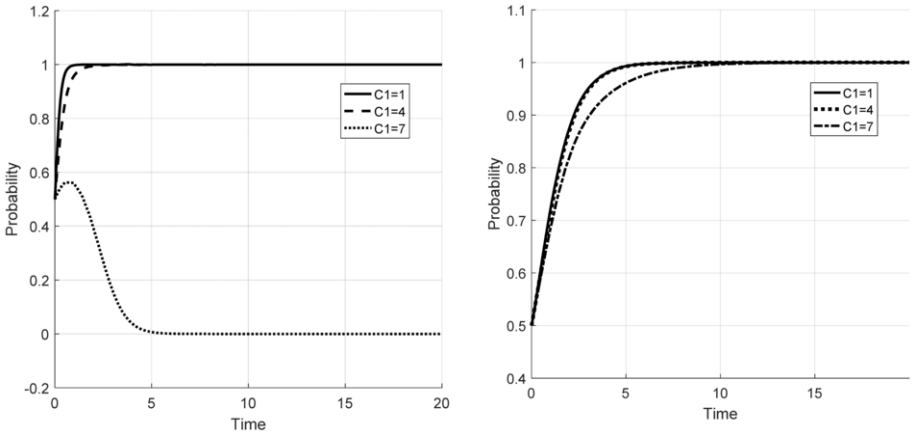
4.2 Impact of parameter changes on evolutionary outcomes

Based on the ideal stable equilibrium point (1,1), this paper analyzes the impact of parameter changes in the game model on the evolutionary outcomes. The initial willingness of the lessor to choose strict supervision and that of the lessee to choose active maintenance are both set to 0.5.

4.2.1 Impact of the lessor's strict supervision cost on the evolutionary outcomes

Set R_1=2, C_2=2, F=3, s=0.5, B=3, t=0.5, H_1=2, and H_2=2. Assume that C_1 takes the values of 1, 4, and 7, respectively. Through numerical simulation, the evolutionary trends of the lessor and lessee under different costs of strict supervision by the lessor are shown in Figure 5. As can be seen from Figure 5, when the lessor's strict supervision cost is relatively low, the lessor chooses the strict supervision strategy. However, as the strict supervision cost increases, the evolutionary speed at which the lessor chooses the strict supervision strategy slows down, and the willingness to conduct strict supervision decreases. Until the strict supervision cost becomes sufficiently high, the willingness to conduct strict supervision weakens, while the willingness to choose the loose supervision strategy strengthens, and the lessor tends to choose the loose supervision strategy. On

the other hand, since the lessee's active maintenance cost remains unchanged, an increase in the lessor's strict supervision cost does not affect the lessee's willingness to choose the active maintenance strategy. However, it is worth noting that as the lessor's strict supervision cost increases, the evolutionary speed at which the lessee chooses the active maintenance strategy slows down. This is because an increase in the lessor's strict supervision cost reduces the lessor's willingness to choose strict supervision, which leads to relaxed supervision of the lessee and a loss of incentive for strict supervision. When the lessor tends to adopt loose supervision, the lessee is fully aware of this, and its willingness to actively choose the active maintenance strategy decreases, thereby slowing down the evolutionary speed at which the lessee actively chooses the active maintenance strategy.



(a) Impact of strict supervision cost on the lessor's evolutionary outcome.

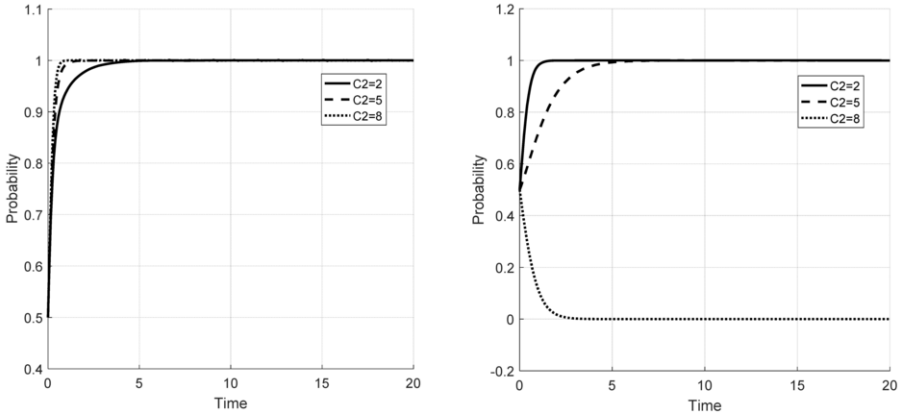
(b) Impact of strict supervision cost on the lessee's evolutionary outcome.

Figure 5. Impact of the lessor's strict supervision cost on the evolutionary outcomes.

4.2.2 Impact of the lessee's active maintenance cost on the evolutionary outcomes

Set R_1=2, C_1=1, F=3, s=0.5, B=6, t=0.5, H_1=2, and H_2=5. Assume that C_2 takes the values of 2, 5, and 8, respectively. Through numerical simulation, the evolutionary trends of the lessor and lessee under different costs of active maintenance by the lessee are shown in Figure 6. As can be seen from Figure 6, when the lessee's active maintenance cost is relatively low, the lessee is more willing to choose the active maintenance strategy. However, as the active maintenance cost increases, the evolutionary speed at which the lessee chooses the active maintenance strategy slows down, and the willingness to conduct active maintenance decreases. Until the active maintenance cost becomes sufficiently high, the lessee tends to choose the passive maintenance strategy. On the other hand, since the lessor's strict supervision cost remains unchanged, the lessee's active maintenance cost does not affect the lessor's willingness to choose strict supervision. However, it is worth noting that as the lessee's active maintenance cost increases, the evolutionary speed at which the lessor chooses the strict supervision strategy becomes faster. This is because an increase in the lessee's active maintenance cost reduces the lessee's willingness to choose active maintenance, which makes the

lessor's choice of the strict supervision strategy more profitable. Strict supervision by the lessor enables the lessor to impose penalties on a lessee that adopts passive maintenance, thereby increasing the lessor's payoff. Ultimately, this strengthens the lessor's willingness to choose the strict supervision strategy, thereby accelerating the evolutionary speed at which the lessor chooses the strict supervision strategy.



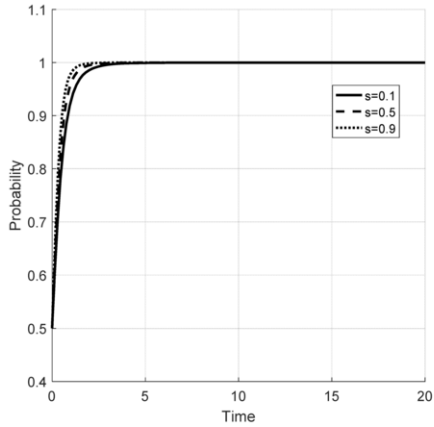
(a) Impact of the active maintenance cost on the lessor's evolutionary outcome.

(b) Impact of the active maintenance cost on the lessee's evolutionary outcome.

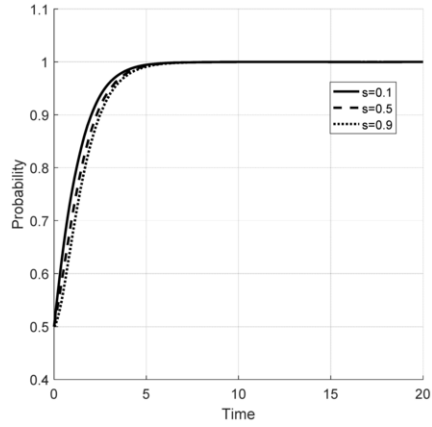
Figure 6. Impact of the lessee's active maintenance cost on evolutionary outcomes.

4.2.3 Impact of the allocation of liability for aircraft accidents on evolutionary outcomes

Set R_1=2, C_1=1, C_2=2, F=3, B=3, t=0.5, H_1=2, and H_2=2. Assume that s takes the values of 0.1, 0.5, and 0.9, respectively. Through numerical simulation, the evolutionary trends of the lessor and lessee under different allocations of aircraft accident liability are shown in Figure 7. As can be seen from Figure 7, when the lessor's share of liability for aircraft accidents is relatively small, that is, when the lessor bears relatively light responsibility in an aircraft accident, its accident compensation amount is relatively small. A relatively small compensation amount cannot motivate the lessor to choose the strict supervision strategy; instead, the lessor is more willing to choose the loose supervision strategy. Therefore, the evolutionary speed at which the lessor chooses the strict supervision strategy shows a slowing trend. In contrast, because the lessee bears a relatively heavy share of liability for aircraft accidents, its accident compensation amount is relatively large. The larger compensation amount exerts a deterrent effect on the lessee, making the lessee more willing to choose the active maintenance strategy. Therefore, the evolutionary speed at which the lessee chooses active maintenance shows an accelerating trend.



(a) Impact of the allocation of liability for aircraft accidents on the lessor's evolutionary outcome.

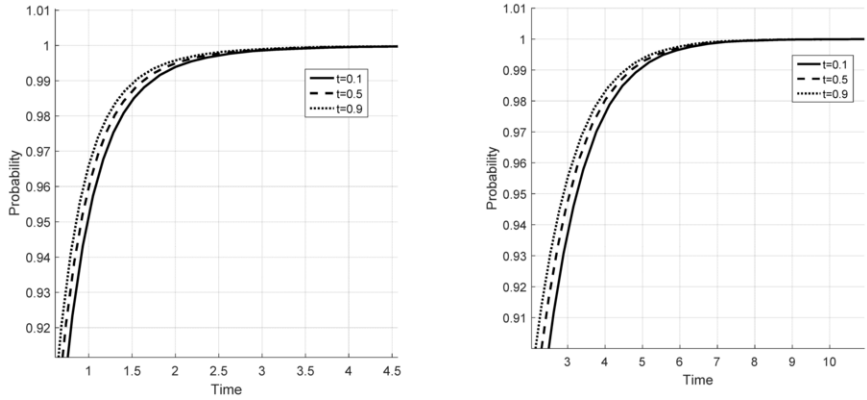


(b) Impact of the allocation of liability for aircraft accidents on the lessee's evolutionary outcome.

Figure 7. Impact of the allocation of liability for aircraft accidents on evolutionary outcomes.

4.2.4 Impact of the probability of public discovery or media exposure of accidents on the evolutionary outcomes

Set R_1=2, C_1=1, C_2=2, F=3, s=0.5, B=3, H_1=2, and H_2=2. Assume that t takes the values of 0.1, 0.5, and 0.9, respectively. Through numerical simulation, the evolutionary trends of the lessor and lessee under different probabilities of public discovery or media exposure of accidents are shown in Figure 8. As can be seen from Figure 8, when the probability that accidents are discovered by the public or exposed by the media is relatively small, the losses suffered by the lessor are relatively small, and the lessor is more willing to choose the loose supervision strategy while the lessee is more willing to choose the active maintenance strategy. However, as the probability that accidents are discovered by the public or exposed by the media increases, the damage to corporate image and social influence caused by public discovery or media exposure becomes greater. Therefore, the lessor's willingness to choose the strict supervision strategy and the lessee's willingness to choose the active maintenance strategy increase, which accelerates the evolutionary speed at which the two parties respectively choose the strict supervision strategy and the active maintenance strategy, and promotes the evolution of the system toward the ideal stable equilibrium point (1,1).



(a) Impact of the probability of public discovery or media exposure of accidents on the lessor's evolutionary outcome. (b) Impact of the probability of public discovery or media exposure of accidents on the lessee's evolutionary outcome.

Figure 8. Impact of the probability of public discovery or media exposure of accidents on the evolutionary outcomes.

5. Conclusions

In the context of developing the low-altitude economy, this paper constructs an evolutionary game model of aircraft leasing composed of aircraft lessors and lessees, and conducts evolutionary stable strategy analysis and numerical simulation. The simulation results show that the evolutionarily stable strategies of the system are the equilibrium points (0,0), (0,1), (1,0), and (1,1), respectively. These equilibrium points correspond to three different stages in the evolution of aircraft leasing strategies:

(1) The embryonic stage of the aircraft leasing market. When the system evolves to the equilibrium point (0,0), the corresponding strategy combination is (loose supervision, passive maintenance), which corresponds to the embryonic stage of the aircraft leasing market. Because the leasing market scale in the embryonic stage is relatively small, aircraft lessors tend to choose the loose supervision strategy, which attracts lessees to enter the aircraft leasing market. At the same time, because lessees lack confidence in the long-term development of the leasing market, or are motivated by short-term profit maximization, they choose the passive maintenance strategy, for example, underinvesting in routine inspection, component replacement, and technological upgrading of aircraft assets, and even concealing potential safety hazards to cut costs.

(2) The growth stage of the aircraft leasing market. The growth stage can be divided into two different paths. The first path is when the system evolves to the equilibrium point (0,1), with the strategy combination (loose supervision, active maintenance). Because the lessor, based on considerations of its own development or policy path dependence, tends to choose the loose supervision strategy, but the lessee, due to increased confidence in the long-term development of the market, begins to actively choose the active maintenance strategy, actively implements aircraft maintenance standards, and improves aircraft reliability. The second path is when the system evolves to the equilibrium point (1,0), with the strategy combination (strict supervision, passive maintenance). Because the lessee has long adapted to a loose supervision environment, it tends to choose the passive maintenance strategy and finds it difficult to adjust its operating model in the short term, neglecting regular maintenance of the aircraft. The lessor realizes this, and strengthening penalties for lessees that adopt passive maintenance can increase the lessor's payoff; therefore, the lessor tends to choose the strict supervision strategy.

(3) The mature stage of the aircraft leasing market. When the system evolves to the equilibrium point (1,1), the corresponding strategy combination is (strict supervision, active maintenance), which corresponds to the mature stage of the aircraft leasing market. At this stage, both lessors and lessees attach importance to the long-term development of the aircraft leasing market. Both parties realize that only by choosing the strict supervision strategy and the active maintenance strategy can they maximize their respective interests and promote the steady and sustainable development of the aircraft leasing market. Therefore, lessors will conduct regular and irregular inspections of the lessee's aircraft maintenance, and introduce emerging technologies such as artificial intelligence to supervise the lessee's aircraft maintenance processes. Lessees will also improve the quality of maintenance and servicing, strengthen the inspection of key aircraft components, enhance the training of maintenance personnel, and engage qualified third parties to participate in the maintenance process, thereby promoting the high-quality development of the aircraft leasing market.

On this basis, this paper analyzes the factors affecting changes in evolutionary strategies. The study finds that the cost of strict supervision by the lessor, the cost of active maintenance by the lessee, the allocation of liability for aircraft accidents, and the probability that accidents are discovered by the public or exposed by the media are the key factors affecting the evolutionary outcomes.

To promote the high-quality development of the aircraft leasing market, the following countermeasures and measures can be adopted:

(1) Lessors should effectively reduce the cost of strict supervision. Lessors should formulate differentiated maintenance standards according to aircraft models and service years, and require lessees to implement them accordingly. In addition, lessors can monitor the operating status of aircraft through technologies such as artificial intelligence, the Internet of Things, and blockchain, strengthen the protection of the lessor's supervision rights and interests in leasing contracts, and avoid repeated supervision caused by information asymmetry, thereby effectively reducing the cost of strict supervision.

(2) Lessees should effectively reduce the cost of active maintenance. Lessees can cooperate with professional third-party institutions on aircraft maintenance, use big data and artificial intelligence technologies for analysis, monitor aircraft operation data in real time, and schedule maintenance in advance to reduce extra costs caused by unexpected failures.

(3) Lessors and lessees need to clarify the allocation of liability for aircraft accidents. The rights and obligations of both parties should be stipulated in the lease contract, the standards for determining liability for aircraft accidents and the methods for dispute resolution should be clarified, aviation liability insurance should be introduced, and timely written records should be kept for the supervision and maintenance processes of both parties.

(4) Strengthen the supervisory role of public discovery and media exposure in the aircraft leasing market. Establish and improve an exposure reward system, give full play to the media's role in public opinion supervision, carry out aviation safety publicity and education, build a multi-stakeholder information release and sharing platform for aircraft leasing, and provide legal protection for public discovery or media exposure.

References

[1] K. Marintseva and R. Athousaki, "Assessment of aircraft leasing efficiency: An airline perspective," Journal of the Air Transport Research Society, vol. 3: 100039, December 2024.

[2] Z. Sun, Y. Wu, L. Zhang, Y. Yang, and X. Chen, "A Comparative Study on Aircraft Leasing Policies Between Tianjin City and Shanghai City," Scientific and Social Research, vol. 6, no. 2, March 2024.

[3] S. Wandelt, X. Sun, and A. Zhang, "Is the aircraft leasing industry on the way to a perfect storm? Finding answers through a literature review and a discussion of challenges," Journal of Air Transport Management, vol. 111: 102426, August 2023.
[4] A. A. Burdina, N. V. Moskvicheva, and N. O. Melik-Aslanova, "Determining the Lease Payments for High-Tech Products," Russian Engineering Research, vol. 42, pp. 619-622, July 2022.
[5] W. Lin, J. Lu, J. Zhu, and L. Xu, "Research on the Sustainable Development and Dynamic Capabilities of China's Aircraft Leasing Industry Based on System Dynamics Theory," Sustainability, vol. 14, no. 3: 1806, February 2022.
[6] G. Dissanaike, R. Jayasekera, and G. Meeks, "Why Do Unsuccessful Companies Survive? U.S. Airlines, Aircraft Leasing, and GE, 2000–2008," Business History Review, vol. 96, no. 3, pp. 615-642, November 2022.
[7] P.Kuhle, D. Arroyo, and E. Schuster, "Building A blockchain-based decentralized digital asset management system for commercial aircraft leasing," Computers in Industry, vol. 126: 103393, April 2021.
[8] S. Bourjade, R. Huc, and C. Muller-Vibes, "Leasing and profitability: Empirical evidence from the airline industry," Transportation Research Part A: Policy and Practice, vol. 97, pp. 30-46, March 2017.
[9] P. Li, J. Chen, H. Yang, and Z. Lin, "Peer-to-peer power trading and pricing for rental energy storage shared community microgrid: A coordinated Stackelberg and cooperative game," Renewable Energy, vol. 256, Part B: 123963, July 2025.
[10] X. Zhu, G. Zhao, G. Yan, J. Li, H. Dong, and C. Li, "Stackelberg Game for Bilateral Transactions between Energy Storage and Wind Farms Considering the Day-Ahead Electricity Market," Energy Engineering, vol. 122, no. 5, pp. 1645–1668, April 2025.
[11] J. Qiao, Y. Mi, S. Ma, Y. Han, and P. Wang, "Trilayer stackelberg game scheduling of active distribution network based on microgrid group leasing shared energy storage," Applied Energy, vol. 382: 125157, March 2025.
[12] B. Zhang and J. Huang, "Shared Energy Storage Capacity Configuration of a Distribution Network System with Multiple Microgrids Based on a Stackelberg Game," Energies, vol. 17, no. 13: 3104, June 2024.
[13] J. Li, Y. Zhu, Y. Xiao, and X. Lan, "Optimized configuration and operation model and economic analysis of shared energy storage based on master-slave game considering load characteristics of PV communities," Journal of Energy Storage, vol. 76: 109841, January 2024.
[14] G. Li and T. Wang, "Long-Term Leases vs. One-Off Purchases: Game Analysis on Battery Swapping Mode Considering Cascade Utilization and Power Structure," Sustainability, vol. 14, no. 24: 16957, December 2022.
[15] L. Zhu, J. Wang, A. Farnoosh, and X. Pan, "A game-theory analysis of electric vehicle adoption in Beijing under license plate control policy," Energy, vol. 244, Part A: 122628, April 2022.
[16] L. Hong, R. Wang, H. Chen, W. Cui, N. Tsoulakos, and R. Yan, "Evolutionary game-based ship inspection planning considering ship competitive interactions," Transportation Research Part E: Logistics and Transportation Review, vol. 196: 103994, April 2025.
[17] F. Ahmad, Zubair Shah, and L. Al-Fagih, "Applications of evolutionary game theory in urban road transport network: A state of the art review," Sustainable Cities and Society, vol. 98: 104791, November 2023.
[18] Z. Liu, N. Huang, B. Hu, Y. Zhao, and I. S. Kown, "An evolutionary game study of traffic safety risk regulation considering public participation behavior," Scientific reports, vol. 14, no. 1: 30069, December 2024.
[19] W. Wu, Z. Lin, and M. Wei, "Research on a collaborative evolution model of multi-airport route network considering subsidy strategies," Mathematical biosciences and engineering, vol. 20, no. 11, pp. 19808-19838, October 2023.
[20] S. Deng, D. Zhou, G. Wu, L. Wang, and G. You, "Evolutionary game analysis of three parties in logistics platforms and freight transportation companies' behavioral strategies for horizontal collaboration considering vehicle capacity utilization," Complex & Intelligent Systems, vol. 9, pp. 1617-1637, September 2022.

Digitalization and Management Innovation V
A. Lau and A.J. Tallón-Ballesteros (Eds.)
© 2026 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).

Author Index

Ait Oufkir, A. 240
Alice, A. 373
Almoqbel, M. 278
An, Y. 350
Bai, S. 35
Bao, Z. 771
Belgnaoui, L.T. 240
Bello, R.-W. 838, 848
Canonigo, A.M. 231
Cao, J. 321
Cao, Y. 675
Chen, G. 515
Chen, W. 301
Chen, X. 321
Chen, Z. 48, 373, 648
Cheng, P. 475
Du, J. 180
Du, T. 321
Efremov, L. 249
Emelianov, S. 76
Fan, B. 468
Fan, D. 858
Fang, W. 901
Feng, B. 434
Galabov, M. 97, 662
Gan, Q. 718
Gao, J. 434
Gao, M. 424
Geng, J. 614
Gong, X. 700
Gu, Y. 373
Guo, F. 689
Guo, J. 373
Guo, Y. 434
Han, H. 871
He, S. 321
He, Z. 820
Hong, Y. 888
Hou, L. 218
Hu, B. 700
Hu, D. 778, 791
Hu, S. 485
Hu, Y.-j. 603
Hu, Z. 485, 730
Huang, H. 895
Huang, J. 154
Huang, X. 130, 547
Huang, Z. 538
Hui, T. 187
Imoru, O. 838
Irmawan, F.N. 350
Jia, X. 593
Jiang, T. 744
Jin, H. 86
Jin, Y. 648
Ke, S. 180
Komarov, M. 64
Korostelev, A. 64
Lafou, T. 240
Lahbabi, Y. 240
Lau, A.S.M. v, 321, 350, 373, 416, 424, 434, 448
Le, Z. 758, 778, 791
Lee, C.-S. 264
Lei, J. 130
Lei, L. 350
Lei, Q. 162
Lei, T. 803
Lei, Z. 434
Li, F. 829
Li, G. 170
Li, J. 523, 718
Li, M. 810
Li, R. 290
Li, S. 130, 416, 424
Li, T. 196, 203, 210
Li, W. 523, 523
Li, X. 180, 881
Li, Y. 170, 623
Liang, S. 778, 791, 810
Liang, Y. 154
Liao, B. 642
Liao, X. 475
Lin, M. 187
Lin, Y. 496
Liu, B. 48

Liu, G. 642, 751
Liu, H. 751
Liu, J. 531, 557, 642, 858
Liu, Q. 48
Liu, S. 858
Liu, X. 523, 871
Liu, Y. 24, 24, 112, 180
Liu, Z. 35
Look, C.-M. 264
Lu, H. 700
Lu, Y. 112, 448
Luo, H. 810
Luo, P. 373
Lv, M. 888
Lyu, P. 434
Ma, L. 290
Ma, Y. 180
Makhorole, T.S. 848
Meng, H. 751
Min, S. 803
Ogundokun, R.O. 838
Owolawi, P.A. 838, 848
Pang, M. 675
Pei, Y. 258
Peng, Y. 48, 508
Pichetjamroen, S. 311
Qian, J. 689
Qin, C. 771
Qiu, G. 642
Qiu, J. 468, 531
Rattanalerdnusorn, E. 311
Ruan, H. 475
Sheng, Z. 810
Shi, M. 424
Song, A. 586
Song, H. 180
Song, J. 142
Su, Y. 829
Sun, B. 196
Sun, J. 35, 187, 538, 744
Tallón-Ballesteros, A.J. v
Tan, J. 218
Tan, M. 475
Tang, G. 203
Tang, H. 858
Tang, Y. 614
Tian, Y. 162, 718
Tu, C. 838, 848
van Wyk, E.A. 838
Velev, D. 97, 400, 662
Volkov, S. 350
Wang, B. 321, 689
Wang, D. 508
Wang, H. 508
Wang, J. 86, 771
Wang, M. 515
Wang, X. 154, 515
Wang, Y. 2, 15, 196, 290, 434, 881
Wang, Z. 642, 744
Wei, J. 858
Wen, J. 258
Wen, Y. 858
Wu, C. 187, 538
Wu, D. 350
Wu, H. 130
Wu, K. 758, 778, 791
Wu, Q. 15, 803
Wu, W. 130
Wu, Y. 123
Wu, Z. 434, 468
Xia, H. 632
Xia, Z. 689
Xiao, T. 290
Xie, L. 210
Xie, R. 424
Yan, H. 758, 778, 791
Yan, X. 632
Yang, A. 523
Yang, G. 301
Yang, H. 515
Yang, L. 895
Yang, W. 180
Yang, Y. 623, 744, 758, 778, 791
Yang, Z. 758
Ye, H. 778, 791
Yin, Z. 424
Yu, B. 895
Yu, Y. 203, 210
Yuan, S. 751
Zeng, L. 468
Zeng, W. 771
Zhang, D. 258
Zhang, H. 895
Zhang, J. 424
Zhang, K. 321
Zhang, L. 881
Zhang, P. 350
Zhang, S. 547

Zhang, W. 778, 791
Zhang, X. 632, 820
Zhang, Y. 48, 301
Zhang, Z. 523, 593
Zheng, S. 154, 803
Zheng, Y. 531
Zhou, H. 258, 575
Zhou, J. 771
Zhou, X. 373
Zhu, Q. 718
Zhu, Y. 24, 112
Zlateva, P. 97, 400, 662
Zou, C. 468, 531
Zuo, K. 350