



Graph Reasoning Over Enterprise Innovation Networks for Industrial Investment Risk Assessment

Emily Carter¹, James Bennett^{2*}

Department of Industrial Engineering, Imperial College London, London SW7 2AZ, United Kingdom

Corresponding Author*: James Bennett E-mail: j.bennett@warwick.ac.uk

ARTICLE INFO

Keywords

Industrial investment risk
enterprise innovation network
knowledge graph reasoning
graph neural network
regional industrial fund patent
commercialization
supply-chain dependency

Published:

03 September 2026

ABSTRACT

Regional industrial funds invest in technology-oriented manufacturing firms, advanced materials companies, intelligent equipment producers, energy enterprises, and digital infrastructure providers. Investment risk is influenced not only by financial indicators but also by patent quality, R&D collaboration, government subsidies, customer concentration, supply-chain stability, executive experience, and related-party transactions. This study proposes a graph reasoning model over enterprise innovation networks for industrial investment risk assessment. The method builds a knowledge graph linking target firms, shareholders, patents, research partners, suppliers, customers, subsidy programs, financing rounds, executives, litigation events, and revenue records. A graph neural network is used to represent innovation capability and business dependency, while knowledge inference identifies hidden investment risks from weak patent commercialization, subsidy dependency, unstable key customers, supplier disruption, and shareholder debt exposure. Experiments are conducted on a regional industrial investment dataset containing 12,800 candidate enterprises, 96,000 patents, 38,400 R&D collaboration links, 520,000 supplier-customer edges, 18,600 subsidy records, 24,300 financing events, and 3,940 post-investment risk cases over 54 months. The proposed model shortens median risk-identification time from 132 days to 57 days after investment screening compared with a due-diligence checklist baseline. It discovers 4,260 innovation-commercialization risk paths and 1,780 subsidy-dependency risk chains. Portfolio simulation shows that graph-informed screening reduces expected impaired investment exposure by 118 million RMB during the validation period. Full quarterly assessment of candidate enterprises is completed in 8.7 minutes. The results demonstrate that enterprise innovation network reasoning can provide interpretable support for industrial fund investment risk assessment.

1. Introduction

Regional industrial funds have become an important policy and financial instrument for promoting technology-oriented manufacturing, advanced materials, intelligent equipment, energy industries, and digital infrastructure. By directing capital toward strategically important enterprises, these funds can accelerate industrial upgrading, technological commercialization, and regional economic development. However, investment decisions in technology-intensive enterprises involve substantial uncertainty because firm value is influenced not only by current financial performance but also by innovation capability, commercialization potential, supply-chain relationships, governance quality, and external financing conditions. Patent value, R&D cooperation, subsidy

Carter, E., & Bennett, J. (2026). Graph Reasoning Over Enterprise Innovation Networks for Industrial Investment Risk Assessment. *The Journal of Applied Engineering and Technologies*, 1(3), 113-121.

3154-7877/© The Authors. Published by J&L Academic Group PLT. This is an open access article under the CC BY 4.0 license.
<https://doi.org/10.64744/tjaet.2026.311>

reliance, customer concentration, supplier stability, executive background, financing history, litigation records, and related-party transactions can all affect the operational sustainability and post-investment performance of portfolio firms (Kaur & Baheti, 2026; You, 2026). These factors are often distributed across different data sources and interact with each other, making industrial investment risk difficult to evaluate through conventional financial indicators alone. Recent research has also emphasized the importance of interpretable analytical models in risk-sensitive decision environments. For example, explainable ensemble learning combined with SHAP has been introduced into network anomaly detection to provide feature-level explanations for model decisions (Wang et al., 2026). Such studies indicate that predictive capability alone is insufficient in practical risk management, where decision makers also need to understand the factors and relationships underlying a model's judgment (Jiao et al., 2026; Y. Wu & Su, 2026).

The increasing availability of enterprise registration records, patent databases, supply-chain information, financing events, judicial records, and public subsidy data has created new opportunities for more comprehensive enterprise risk assessment. Existing studies have used enterprise networks, patent maps, supply-chain relationships, and knowledge graphs to characterize corporate connections, technological capabilities, and innovation activities (J. Wu et al., 2026b; Sabet & Mohammadi, 2026). Compared with traditional tabular representations, network-based approaches can explicitly describe relationships among firms, shareholders, research institutions, suppliers, customers, executives, investors, and government programs. These relationships are particularly important for regional industrial funds because investment risk may originate outside the focal enterprise itself (Lin et al., 2020). Financial distress of a major customer, instability of a key supplier, excessive shareholder leverage, repeated financing dependence, or weak cooperation between patents and commercial products may gradually affect firm performance. Graph neural networks have therefore been increasingly applied to credit scoring, default prediction, supply-chain risk assessment, and firm behavior analysis because they can aggregate information from connected entities and identify patterns that are difficult to capture from isolated enterprise records (Bao & Wang, 2026b; Mojdehi et al., 2025). These developments provide an important methodological foundation for modeling investment risk as a relational rather than purely firm-level problem. Despite these advances, current enterprise risk assessment methods remain insufficient for regional industrial fund investment. Many studies continue to rely primarily on financial ratios, patent counts, credit records, or a single category of relational data (Chen et al., 2025; Qi & Qiao, 2026). Such representations cannot adequately distinguish firms with similar accounting performance but substantially different innovation quality, commercialization ability, customer stability, subsidy dependence, or shareholder financial exposure (Valaskova et al., 2025; Xu et al., 2026). Patent quantity, for example, does not necessarily indicate that patented technologies have been successfully transformed into products, contracts, or stable revenue (Moro-Visconti, 2025; Yin et al., 2024). Likewise, rapid revenue growth may conceal excessive dependence on a small number of customers or government subsidies. These risks become more difficult to identify when relevant evidence is dispersed across patent ownership, research cooperation, procurement relationships, financing records, litigation events, and related-party transactions. Another limitation concerns model interpretability. Many machine learning and graph-based approaches are developed primarily to improve classification or prediction accuracy, while providing limited information about the specific relationships that produce a high-risk judgment (Hong et al., 2026; Xiong et al., 2026). This limitation is particularly important in industrial investment due diligence, where investment committees and fund managers need to understand why an enterprise is classified as risky and which underlying factors require further verification (Li

et al., 2026; Plass & Robinson, 2025). A numerical risk score alone provides limited support for decisions involving capital allocation, due diligence priorities, investment conditions, or post-investment monitoring. An enterprise may be identified as high risk because of weak patent commercialization, dependence on a single customer, repeated subsidy-supported revenue, unstable upstream suppliers, shareholder debt exposure, or interconnected litigation events. Distinguishing among these risk mechanisms is therefore necessary for transforming predictive models into practical decision-support tools (J. Wu et al., 2026a; Michou et al., 2026). The characteristics of available data create an additional challenge. Experimental studies frequently use relatively small datasets or samples dominated by publicly listed firms, whose financial statements, ownership information, and operating records are comparatively complete (Huang et al., 2026; Zheng & Makar, 2022). Regional industrial funds, in contrast, often evaluate large numbers of private technology enterprises at different stages of development. These firms may have limited financial history, incomplete disclosure, rapidly changing ownership structures, irregular financing records, and heterogeneous operational information. Important signals may appear in nonfinancial sources long before deterioration becomes visible in financial statements. Effective industrial investment assessment therefore requires a framework capable of integrating heterogeneous enterprise information, learning from interconnected entities, and identifying meaningful risk relationships even when individual firm records are incomplete.

To address these problems, this study develops a graph reasoning model over enterprise innovation networks for industrial investment risk assessment. The proposed framework constructs a heterogeneous knowledge graph connecting enterprises with shareholders, patents, research partners, suppliers, customers, subsidy programs, financing rounds, executives, litigation events, and revenue records. Graph representation learning is used to capture structural dependencies and latent risk information across interconnected entities, while rule-based reasoning is introduced to identify interpretable risk paths associated with patent commercialization, subsidy dependence, customer instability, supplier disruption, and shareholder debt exposure. By combining relational representation learning with explicit reasoning, the model is designed not only to identify enterprises with elevated investment risk but also to reveal the network structures and event chains contributing to that assessment. This approach extends industrial investment evaluation from isolated financial scoring to multidimensional enterprise-network analysis and provides a more suitable framework for technology-intensive private firms with heterogeneous and incomplete records. The study aims to improve early risk identification, strengthen the interpretability of investment screening, and support due diligence, portfolio selection, and post-investment monitoring for regional industrial funds. More broadly, it provides a scalable approach for transforming fragmented enterprise information into structured and explainable evidence that can support investment decisions in technology-driven regional industrial development.

2. Materials and Methods

2.1 Sample and Study Area Description

This study examined regional industrial fund investment in technology-focused enterprises, including advanced manufacturing, new materials, intelligent equipment, energy, and digital infrastructure firms. The sample included 12,800 enterprises collected over 54 months from a regional industrial investment database. Each enterprise had records for patents, financing events, subsidies, supply-chain relationships, shareholders, executives, litigation cases, and post-investment outcomes. The dataset contained 96,000 patents, 38,400 R&D collaboration links,

520,000 supplier-customer connections, 18,600 subsidy records, 24,300 financing events, and 3,940 confirmed post-investment risk cases. Enterprises ranged from early-stage startups to mature firms, allowing assessment of risks across different development stages. This sample reflects the practical conditions faced by regional industrial funds, which evaluate firms with varied financial health, innovation capacity, and policy dependence.

2.2 Experimental Design and Control Experiments

The experiment compared the proposed graph reasoning model with traditional investment screening methods. The experimental group used an enterprise innovation network constructed from multi-source data, while the control group applied a checklist-based screening method considering financial ratios, patent counts, subsidy amounts, litigation records, and customer concentration. Traditional methods treat each enterprise independently, whereas investment risk often spreads through networks of shareholders, suppliers, customers, and partners. Two additional comparison models were included: a regression-based risk score model and a machine learning model using tabular enterprise features. Data were split chronologically into training, validation, and testing sets to prevent information leakage. Enterprises from earlier quarters trained the models, and later quarters tested whether risks could be identified before negative outcomes occurred.

2.3 Measurement Methods and Quality Control

Investment risk was measured using confirmed post-investment risk cases and intermediate indicators. A post-investment risk case was defined as an enterprise showing one or more negative outcomes, such as project delay, revenue loss, legal dispute, subsidy withdrawal, key customer loss, supplier disruption, debt default, or investment impairment. Patent quality was assessed by patent type, legal status, citation count, ownership stability, and linkage with enterprise revenue. R&D collaboration was measured by the number, duration, and type of research partners. Supply-chain dependency was measured by the proportion of critical suppliers and customers. Subsidy dependency was calculated as the share of subsidies in operating income and consistency over time. Data quality control included three steps: standardizing enterprise names and entity records, removing duplicates and inconsistent timestamps, and checking abnormal values against source records. Missing financial data were imputed using industry-year medians, while missing relational data were maintained as sparse connections to avoid introducing false links.

2.4 Data Processing and Model Formulation

Data were transformed into a heterogeneous enterprise innovation graph. Nodes represented entities such as enterprises, shareholders, patents, executives, suppliers, customers, subsidies, financing events, and litigation cases. Edges represented verified relationships, including ownership, patent holding, R&D collaboration, supply transactions, customer links, subsidy receipt, financing participation, executive affiliation, and litigation involvement. Numerical features were normalized by industry and year, and categorical features were one-hot encoded or embedded. A baseline regression model estimated investment risk:

$$\text{Risk}_i = \beta_0 + \beta_1 \text{Fin}_i + \beta_2 \text{Pat}_i + \beta_3 \text{Sub}_i + \beta_4 \text{Sup}_i + \beta_5 \text{Lit}_i + \varepsilon_i$$

where Fin_i is financial indicators, Pat_i patent measures, Sub_i subsidy reliance, Sup_i supply-chain concentration, and Lit_i litigation exposure. In the graph model, node representations were updated as:

$$h_i^{(l+1)} = \sigma \left(w^{(l)} h_i^{(l)} + \sum_{j \in N(i)} \alpha_{ij}^{(l)} w_r^{(l)} h_j^{(l)} \right)$$

where $h_i^{(l)}$ is the representation of node i at layer l , $N(i)$ is the set of neighbors, $\alpha_{ij}^{(l)}$ is the edge weight, $w^{(l)}$ and $w_r^{(l)}$ are trainable parameters, and $\sigma(\cdot)$ is a nonlinear function. The resulting representation was used to predict investment risk and identify interpretable risk paths.

2.5 Model Evaluation and Implementation

Model performance was evaluated in three aspects: prediction accuracy, early warning capability, and interpretability. Accuracy was measured using AUC, precision, recall, F1-score, and calibration. Early warning efficiency was the median time between risk detection and confirmed post-investment risk. Interpretability was assessed by the number and quality of detected risk paths, including weak patent commercialization, subsidy dependence, unstable customers, supplier disruption, and shareholder debt exposure. Portfolio-level impact was evaluated through simulations comparing impaired investments under graph-informed and checklist-based screening. All experiments used the same quarterly assessment schedule to ensure comparability. The model was implemented with a graph learning framework and mini-batch training to handle the large-scale enterprise network. The complete quarterly assessment for all enterprises was completed in 8.7 minutes, demonstrating practical feasibility for industrial fund screening.

3. Results and Discussion

3.1 Risk Prediction Performance

The graph reasoning model produced better prediction results than the checklist baseline, regression model, and tabular machine learning model. This improvement was mainly related to the use of enterprise relationships rather than isolated firm records. The checklist method used separate indicators, such as financial ratios, patent counts, subsidy amounts, and litigation records. Therefore, it failed to capture indirect risks caused by customers, suppliers, shareholders, and research partners. In contrast, the proposed model learned risk signals from connected entities, including enterprises, patents, suppliers, customers, subsidies, financing events, and litigation cases. As shown in Fig.1, the graph reasoning model showed more stable results across quarterly testing periods, while the baseline methods were more sensitive to incomplete enterprise records. This result is consistent with recent graph-based credit risk research, which shows that graph structures can improve prediction by using links among related records instead of treating each sample as independent (Gu, 2026; Mesinovic et al., 2026). The finding indicates that enterprise investment risk should be evaluated within a relational structure, especially when firms are strongly connected through innovation, supply chains, and financing activities.

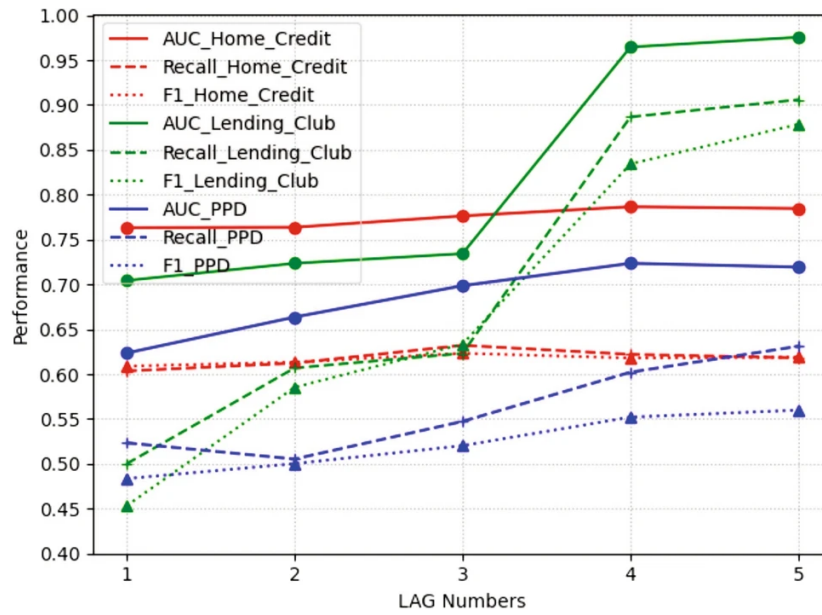


Fig. 1. Predictive performance of the graph reasoning model compared with baseline approaches, showing higher AUC and F1 scores for early enterprise risk detection.

3.2 Early Identification of Post-Investment Risk

The proposed model reduced the median risk-identification time from 132 days to 57 days after investment screening. This result suggests that network-based indicators can reveal risk before clear financial deterioration appears. Some enterprises maintained stable short-term revenue but showed weak patent commercialization, unstable subsidy support, and high dependence on a small number of customers. These firms were difficult to identify through checklist screening because no single indicator reached a severe warning level. However, the graph model connected patent ownership, subsidy history, customer concentration, and financing events, allowing earlier detection of hidden risk paths. Compared with traditional credit prediction methods that mainly focus on feature selection and classifier performance, the proposed method provides earlier warning by tracing changes in enterprise relationships over time (Hezam et al., 2025; Zhao et al., 2026). This advantage is important for regional industrial funds because earlier detection can support follow-up investigation, staged capital release, and portfolio adjustment.

3.3 Risk Path Interpretation and Network Effects

The model identified 4,260 innovation-commercialization risk paths and 1,780 subsidy-dependency risk chains during the validation period. The most common risk pattern was weak patent commercialization combined with high subsidy dependence and unstable customer links. This result indicates that some firms had visible innovation activities but lacked stable market demand. Another important pattern was supplier disruption combined with shareholder debt exposure, which increased the possibility of delayed delivery and investment impairment. Fig.2 shows a representative multi-layer risk path in which patent quality, subsidy reliance, supplier concentration, and customer instability jointly contribute to post-investment risk. This finding is consistent with recent knowledge graph studies on supply-chain decision support, where graph reasoning was used to infer hidden links and improve transparency when data were sparse or incomplete (Bao & Wang, 2026a; Peeris et al., 2026). Compared with studies focused only on supply-chain visibility or sustainability links, this study further combines innovation, financing, legal, and business-operation records. Therefore, the identified risk paths are more suitable for industrial fund screening.

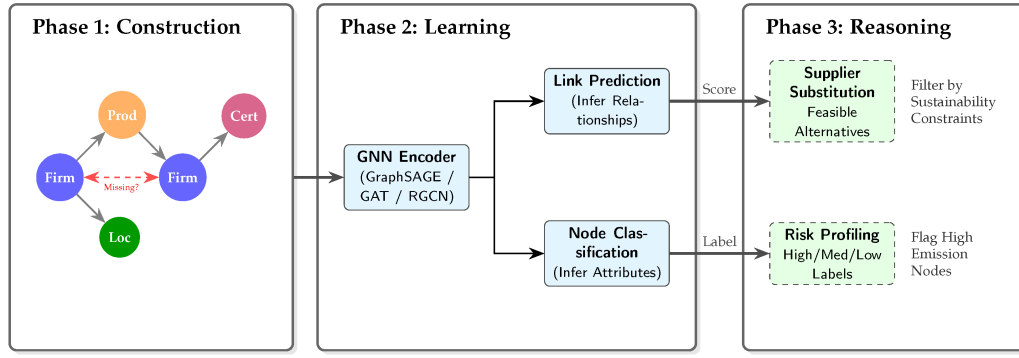


Fig. 2. Example of a multi-layer risk path illustrating the combined effects of patent quality, subsidy reliance, supplier concentration, and customer instability.

3.4 Portfolio Effect and Practical Value

Portfolio simulation showed that graph-informed screening reduced expected impaired investment exposure by 118 million RMB during the validation period. This reduction was mainly achieved by avoiding firms with hidden dependency risks, rather than simply excluding firms with weak financial ratios. The result shows that industrial investment risk is not only a firm-level financial issue but also a network-level issue shaped by innovation partners, customers, suppliers, subsidies, shareholders, and financing relationships. Compared with previous graph learning studies on loan default prediction, this study covers a broader industrial investment setting and includes both innovation-related and operation-related entities. The quarterly assessment of all candidate enterprises was completed in 8.7 minutes, showing that the method can support large-scale preliminary screening. However, the model should be used as a decision-support tool rather than a full replacement for due diligence. For high-value investment cases, graph-based results should be combined with site visits, contract review, technical assessment, and management interviews.

4. Conclusion

This study proposed a graph reasoning model based on enterprise innovation networks for industrial investment risk assessment. The model linked enterprises, shareholders, patents, R&D partners, suppliers, customers, subsidies, financing events, executives, litigation records, and revenue data to detect risk signals that are often missed by traditional checklist screening. The results showed that the proposed model improved prediction performance, reduced the median risk-identification time from 132 days to 57 days, detected 4,260 innovation-commercialization risk paths and 1,780 subsidy-dependency risk chains, and reduced expected impaired investment exposure by 118 million RMB in the portfolio simulation. The main contribution of this study is the joint use of graph representation learning and knowledge reasoning, which supports both risk prediction and risk-path explanation. The findings indicate that industrial investment risk is affected not only by firm-level financial status but also by innovation links, policy dependence, supply-chain structure, and ownership relations. The proposed method can be used for early screening, follow-up due diligence, portfolio monitoring, and post-investment risk control by regional industrial funds. This study also has several limitations. The dataset came from one regional investment setting, and some private enterprise records may include missing or delayed information. Future studies should test the model in more regions and industries, include real-time operating data, and further examine the causal links behind risk transmission.

References

1. Bao, Y., & Wang, H. (2026a). Edge-to-Cloud Infrastructure Continuum: A Unified Framework for Optimizing Industrial IoT Systems. Available at SSRN 7185578.
2. Bao, Y., & Wang, H. (2026b). Large-scale Metrics Migration: A Systematic Approach to Rebuilding Production Measurement Infrastructure. Available at SSRN 7185660.
3. Chen, H., Ning, P., Li, J., & Mao, Y. (2025). Energy Consumption Analysis and Optimization of Speech Algorithms for Intelligent Terminals.
4. Gu, X. (2026). Identifying Causal Effects and Analyzing Heterogeneity of User Growth Interventions on Digital Platforms: Evidence from Large-Scale Behavioral Data. Available at SSRN 6809181.
5. Hezam, Y., Luong, H., & Anthonysamy, L. (2025). Machine learning in predicting firm performance: a systematic review. *China accounting and finance review*, 27(3), 309-339.
6. Hong, Z., Xu, T., & Chen, H. (2026). Automated Detection and Reclamation of Orphaned Compute and Storage Resources in Cloud Infrastructure. Available at SSRN 7116258.
7. Huang, J., Xu, T., Yang, J., & Yin, J. (2026). A Graph Neural Network Approach for Financial Data Validation and Risk Identification in Large-Scale Intercompany Transactions. Available at SSRN 7175058.
8. Jiao, Y., Shi, T., Zhao, B., & Wang, A. (2026). Retrieval-Guided Structured Reasoning and Interpretable Representation Learning for Large-Scale Video–Language Models.
9. Kaur, H., & Baheti, A. (2026). The Silent Shareholder: Understanding the Governance Gap in India’s Retail Investing Boom.
10. Li, J., Ma, R., & Gu, Y. (2026). From Audit Requirements to Computable Software Architecture: A Rule-Engine and Evidence-Indexing Method for Trusted Digital Infrastructure.
11. Lin, N., Xu, Y., Yang, H., Zhang, G., Zhang, M., Wang, S., ... & Li, X. (2020). Dissociating the neural correlates of the sociality and plausibility effects in simple conceptual combination. *Brain Structure and Function*, 225(3), 995-1008.
12. Mesinovic, M., Buhlan, M., & Zhu, T. (2026). Causal graph neural networks for healthcare. *Nature Biomedical Engineering*, 1-21.
13. Michou, I., Maroulis, I., Hatzilygeroudis, I., & Koutsojannis, C. (2026). A Translational Predictive Analytics Framework for Explainable Risk Assessment: Transforming High-Dimensional Surgical Data into Clinical Decision Support Tiers (S-CRI). *Applied Sciences*, 16(13), 6745.
14. Mojdehi, K. F., Amiri, B., & Haddadi, A. (2025). A novel hybrid model for credit risk assessment of supply chain finance based on topological data analysis and graph neural network. *IEEE Access*, 13, 13101-13127.
15. Moro-Visconti, R. (2025). Patent Licensing and Monetization Strategies. In *Patent Valuation: Economic, Financial, and Market Approaches* (pp. 493-535). Cham: Springer Nature Switzerland.

16. Peeris, M. P., Baryannis, G., & Papadakis, E. (2026). Advancing Sustainable Supply Chains Through Knowledge Graph Completion and Graph-Based Artificial Intelligence. *Sustainability*, 18(6), 2825.
17. Plass, L., & Robinson, S. (2025). Physical Climate Risk in the Built Environment: An Institutional Investor View. *Journal of Portfolio Management*, 51(11), 124.
18. Qi, C., & Qiao, X. (2026). Building Reliable AI Systems: A Framework That Bridges Architecture and Operations. Available at SSRN 6795298.
19. Sabet, M., & Mohammadi, N. (2026). Analyzing technological trends in maritime supply chains: A multi-method patent analysis. *Transport Policy*, 104150.
20. Valaskova, K., Nagy, M., & Juracka, D. (2025). Digital transformation and financial performance: An empirical analysis of strategic alignment in the digital age. *Journal of Enterprising Communities: People and Places in the Global Economy*, 19(5), 1178-1205.
21. Wang, S., Feng, Y., & Fang, X. (2026, May). A Large Language Model-Enabled Multi-Agent Collaboration Method for Complex Task Solving. In 2026 6th International Symposium on Computer Technology and Information Science (ISCTIS) (pp. 253-256). IEEE.
22. Wu, J., Wu, D., Zhang, J., & Peng, Y. (2026a). Generative AI Feedback in Junior High School Artistic Creation Evaluation. Available at SSRN 7244838.
23. Wu, J., Zhang, J., Wu, D., & Peng, Y. (2026b). Visual Transformation Mechanisms for Integrating Digital Cultural Heritage Resources into Junior High School Art Curricula.
24. Wu, Y., & Su, D. (2026). An Evaluation of Parental Question Characteristics and the Quality of AI Recommendations in Family Counseling for Children with Autism.
25. Xiong, W., Zeng, Y., & Guo, Y. (2026). A Study on the Characterization of Building MEP System States and the Learning of Behavioral Patterns Driven by Large-Scale Operational Data. Available at SSRN 7121883.
26. Xu, T., Zhu, W., & Zhang, J. (2026). Stability and Consistency of Explainable Deep Learning Methods in Credit Risk Assessment. Available at SSRN 6893938.
27. Yin, L., Tong, Y., Xie, R., Zhang, Z., Islam, Z. H., Zhang, K., ... & Wang, B. (2024). Targeted NAD⁺ repletion via biomimetic nanoparticle enables simultaneous management of intimal hyperplasia and accelerated re-endothelialization: a proof-of-concept study toward next-generation of endothelium-protective, anti-restenotic therapy. *Journal of Controlled Release*, 376, 806-815.
28. You, S. (2026). Verifiable Audit Mechanisms in AI Compliance Automation: Scalability. Available at SSRN 6547458.
29. Zhao, J., Fan, J., & Li, L. (2026). A Study on an Explainable Causal-Enhanced LLM Agent for Predicting the Forming Quality of Automotive Component Materials.
30. Zheng, J., & Makar, M. (2022). Causally motivated multi-shortcut identification and removal. *Advances in Neural Information Processing Systems*, 35, 12800-12812.