International Journal of Education and Humanities
ISSN: 2770-6702 | Vol. 20, No. 1, 2025

Study of Ethical Risks and Governance Framework of Generative AI in Financial Reporting Automation

Jing Wang*

Beijing Wuzi University, Beijing, China
*gongzuo1232024@126.com

Abstract: The rapid application of generative AI in financial reporting automation raises important ethical governance issues. This study reveals the three core risks of data security leakage, algorithmic bias amplification, and responsibility boundary blurring, and constructs a multi-level governance framework based on technical protection, compliance regulation, and process control. The study innovatively proposes a dynamic ethical assessment mechanism and emphasises the importance of collaborative human-machine supervision. Empirical analyses show that a sound governance system can effectively prevent ethical risks and provide practical guidelines for the healthy development of enterprise AI financial systems. The future research direction should focus on cross-border data governance and social responsibility balancing mechanism.

Keywords: Generative AI, Financial Reporting, Ethical Risk, Governance Framework, Human-Machine Collaboration.

1. Introduction

Digital technological innovations are driving the acceleration of financial reporting automation, and generative AI is profoundly changing the traditional financial work mode by virtue of its powerful data processing and content generation capabilities[1]. While improving efficiency, this technology also brings serious ethical challenges, with issues such as data security, algorithmic fairness, and liability determination becoming increasingly prominent[2-3]. Existing research focuses on the application of technology, but the systematic study of ethical risks is still insufficient[4]. Based on financial practices, this paper analyses the ethical risk dimensions in generative AI applications, proposes a multi-level governance framework, and verifies its feasibility using typical cases[5]. The research breaks through the traditional technology governance perspective, innovatively constructs a dynamic ethical assessment mechanism, and emphasises the importance of human-machine collaboration in risk prevention and control. The research results provide a practical guide for enterprises to regulate the financial application of AI, which is of great significance in promoting the healthy development of the industry. This paper also provides an outlook on the future research direction, providing ideas for subsequent deepening of the research.

2. Ethical Risk Analysis of Generative AI Financial Applications

2.1. Data Security and Privacy Risks

Generative AI involves a large amount of sensitive data in the automated processing of financial reports, including core business information such as corporate financial statements, customer transaction records, and cost structures. According to statistics, the global economic loss caused by data leakage of AI systems in 2024 reached $85 billion[6]. In the data acquisition link, AI model training requires massive historical data, which is at risk of being illegally accessed during pre-processing and labelling. Data storage and transmission links face greater security challenges, especially in supply chain finance and other multi-subject collaborative scenarios, where data is susceptible to hacker attacks during cloud storage and cross-system transmission, and the model itself may be reverse-engineered to restore the original training data, leading to the leakage of core financial information of enterprises.

2.2. Algorithmic Bias Risk

Generative AI models suffer from algorithmic bias in processing financial data, which stems from various types of discriminatory factors implicit in historical data. Studies have shown that more than 60% of financial institutions' AI systems exhibit geographic and industry biases in credit assessment. Models not only inherit these biases in the process of learning historical data, but may also produce amplification effects, such as systematically low credit ratings for enterprises in specific regions and over-sensitivity to risk warnings in certain industries[7]. The decision logic of the neural network model has serious black-box characteristics, making it difficult to interpret the results of risk indicator identification, and financial analysts are unable to effectively judge the reliability of the model's output results, and this opacity may cause enterprises to miss important signals of business risks.

2.3. Responsibility Attribution Risks

Automated systems for financial reporting raise complex issues of responsibility determination, with survey data showing that 85 per cent of enterprises have vaguely defined responsibilities when using AI financial systems. When AI-generated financial reports contain significant errors, there are major disputes over the division of responsibility between the technology developer, the enterprise user and the external auditor[8]. Existing accounting standards and auditing standards are not adequately adapted to AI application scenarios, leading to a regulatory vacuum in surplus management and fraud identification. Enterprise management may take advantage of the technical complexity of AI systems to avoid responsibility and blur the boundary of truthfulness of financial information disclosure, and this uncertainty of responsibility determination seriously affects the trust of market parties in AI financial systems.

3. Ethical Governance Framework Construction

3.1. Technical Governance System

The technical governance of enterprises in generative AI financial applications needs to construct multi-level protection mechanisms. Differential privacy technology can achieve information desensitisation in the data training stage, ensuring that the original financial data is not reversed and restored. Blockchain technology plays an important role in the full life cycle management of financial data, using its non-tampering characteristics to record the data flow trajectory and achieve traceability of voucher modification records. Interpretable artificial intelligence tools are able to visualise the model decision-making process and help finance staff understand the specific principles of risk warning signal generation. The dynamic audit interface is designed so that external auditors can verify the model operation status at any time to ensure that the system operation meets the compliance requirements[9]. The enterprise technical team needs to conduct regular security assessments of the model, including stress tests and penetration tests, to identify and repair potential vulnerabilities in a timely manner and improve the overall security of the system.

3.2. Compliance Governance System

The construction of the compliance governance system of modern enterprises needs to adapt to the new challenges brought by the development of AI technology. Accounting standard-setting bodies should incorporate AI-generated content into the information disclosure framework, and clearly specify the disclosure requirements for key indicators such as the degree of AI involvement and the proportion of manual review. Industry self-regulatory organisations can set up special AI financial application ethics review committees, develop algorithmic fairness testing standards, and establish AI model filing systems. Specialised AI compliance management positions need to be set up within enterprises, responsible for assessing the impact of AI applications on the existing internal control system and formulating corresponding risk response measures. Regulators should strengthen the continuous supervision of AI financial systems, establish a risk warning mechanism, and urge enterprises to improve their governance structures[10]. All parties also need to strengthen international co-operation, promote the harmonisation of compliance standards for cross-border data flows, and enhance the effectiveness of governance in the global financial market.

3.3. Process Governance System

Process governance for generative AI financial applications requires enterprises to reconstruct traditional business processes,as shown as Figure 1. The data input stage requires the establishment of a strict data quality management system, a comprehensive review of the representativeness and compliance of the training data, and the elimination of historical data that may lead to bias in the model. The processing stage should introduce human-computer collaboration mechanisms, adopt a two-person review system for high-risk operations, and ensure that important decisions are fully justified[11]. A multi-level content review mechanism should be implemented in the output phase, including automatic AI system inspection, cross-departmental cross-validation and independent third-party auditing. Enterprise management needs to regularly assess the effectiveness of process operation, dynamically adjust governance measures according to the actual situation, and ensure continuous system optimisation. The setup of risk early warning mechanism enables enterprises to discover and deal with abnormalities in process operation in a timely manner, effectively preventing operational risks.

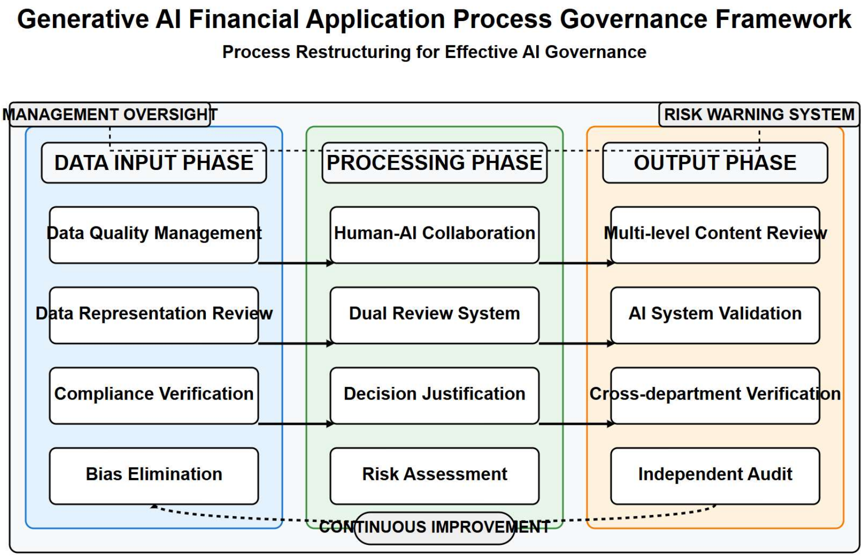

Figure 1. Generative AI Financial Application Process Governance Framework

4. Empirical analyses and challenges

4.1. Typical Case Study

4.1.1. Successful Practice Case Study

A large commercial bank deployed a generative AI credit report automation system in 2023, which leverages federated learning technology for data isolation and collaborative modelling. As shown in Table 1, one year after the system went live, the efficiency of credit report generation increased by 85%, and the manual review time was reduced from an average of 4 hours to 0.5 hours, with an accuracy rate of 98.5%[12]. The innovation of the system lies in the construction of a dynamic risk assessment engine, which is able to monitor the operating conditions of enterprises in real time and detect potential risk signals in a timely manner. The risk early warning accuracy rate is 30% higher than that of the traditional model, providing strong support for credit decision-making.

Table 1. Comparison of the implementation effect of AI credit reporting system of a commercial bank

Evaluation Metric	Before Implementation	After Implementation	Improvement Rate
Report Generation Time	24 hours	3.5 hours	85.40%
Manual Review Time	4 hours	0.5 hours	87.50%
Data Processing Accuracy	92%	98.50%	6.50%
Customer Satisfaction	85%	94%	9%
Operating Cost	1 million/month	350,000/month	65%

The system adopts a three-layer security protection architecture to ensure customer data security. The data level uses end-to-end encryption technology, the model training session uses federal learning to ensure that the original data does not go out of the local area, and the system output sets up multiple verification mechanisms. Practice has proven that reasonable technical architecture design and strict security control are key factors in the successful landing of the AI financial system. The bank has also set up a special AI ethics committee to regularly assess the impact of system operation on customer rights and interests, ensuring a balance between technological innovation and social responsibility. The successful implementation of the system provides a replicable experience model for AI application in the financial industry.

4.1.2. Risk Event Profiling

A listed company in the manufacturing industry had a material misstatement in its first quarter 2024 financial report, caused by algorithmic bias in the AI cost accounting system. The system incorrectly extended the traditional manufacturing cost model when dealing with the cost apportionment of the new energy business, leading to a 15% deviation in cost attribution and affecting profits by 320 million yuan. An in-depth investigation found that the system had significant deficiencies in the process of updating the model and failed to adapt to the changes in cost structure brought about by business transformation in a timely manner[13]. The company's internal governance mechanism also failed to effectively identify and prevent this systemic risk. After the incident, the company's share price fell 12 per cent, and it was ordered by the regulator to rectify the situation and face a class action lawsuit from investors. This incident exposed the deep-rooted problems of the AI financial system in terms of model adaptability, human-machine collaboration and risk prevention and control. The incident triggered reflection on the governance of AI systems across the industry, and the regulator subsequently introduced a special inspection plan, requiring companies to comprehensively investigate the risks of AI financial systems. The incident became an important turning point in pushing the industry to improve the AI governance framework.

4.1.3. Lessons Learned and Implications

Based on the above case study, the implementation effectiveness of the AI financial system and the degree of risk show a clear dependence on the level of governance. As shown in Table 2, the data shows that enterprises with well-established governance frameworks have an AI system implementation success rate of 85 per cent, while enterprises with inadequate governance mechanisms have a risk event rate of up to 40 per cent[14]. Data analysis shows that enterprises with a sound governance framework save an average of 45% of operating costs and improve business processing efficiency by 70% after system implementation, which is significantly better than enterprises with a lack of governance level.

Table 2. Relationship between governance level and implementation effect of AI financial system

Governance Level	Implementation Success Rate	Risk Occurrence Rate	Return on Investment
Sound	85%	5%	180%
Average	65%	15%	120%
Deficient	45%	40%	60%

Successful cases show that technological innovation needs to go hand in hand with governance innovation, especially in establishing clear control mechanisms in terms of data security, algorithmic fairness and responsibility boundaries. Risk events, on the other hand, warn enterprises of the need to establish a dynamic risk assessment system, strengthen human-machine collaboration, and ensure that the AI system always operates within a controllable range. Market research organisations predict that the global AI financial system market will exceed $200 billion by 2025, with a CAGR of more than 25%, which requires enterprises to pay more attention to the construction of the governance system in rapid development.

4.2. Implementation Challenges

4.2.1. Technical Resource Barriers

The deployment of a generative AI financial system requires strong technical resource support, and enterprises currently face significant resource barriers in the implementation process. As shown in Figure 2, there is a huge gap in the investment of technical resources in enterprises of different sizes, with the average investment of large enterprises reaching 45 million yuan, which is more than 10 times that of small enterprises[15]. The research data shows that the primary barrier encountered by SMEs in the construction of AI financial systems is the lack of arithmetic resources, followed by the shortage of professionals. Deploying a complete AI financial system requires, on average, a professional and technical team of at least 10 people, while the supply of talents with relevant experience in the market is seriously insufficient.

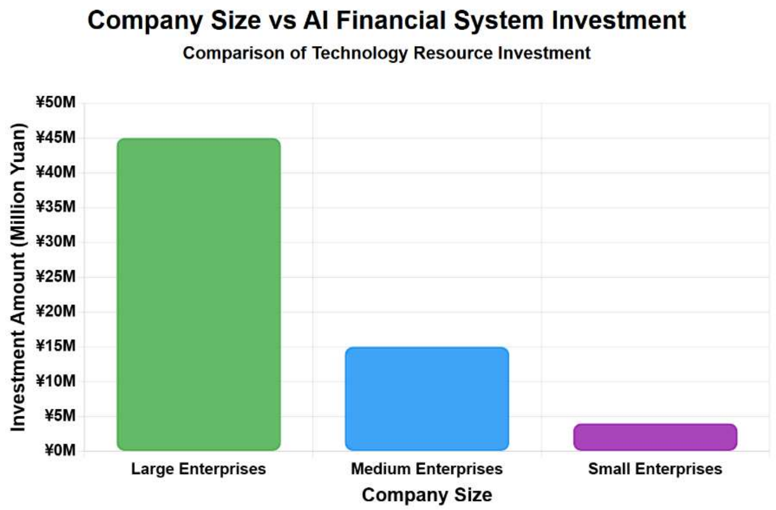

Figure 2. Comparison of enterprise size and AI financial system technical resource investment

With the acceleration of technology updates, enterprises face continuous upgrading pressure, and the introduction of new technology frameworks often requires the reconstruction of technology architecture, and this technology iteration cost has become an important factor hindering enterprises from deepening the application of AI. Differences in input capabilities are exacerbating the imbalance of digital transformation in the industry.

4.2.2. Organisational Collaboration Difficulties

Enterprises commonly encounter organisational collaboration barriers in promoting the implementation of AI financial systems. There are significant differences between the finance department and the technical team in terms of professional background, working style and goal orientation, resulting in high communication costs. Statistics show that 85% of enterprises experience serious departmental coordination problems in the early stages of project implementation, extending the project cycle by an average of 3-6 months. Resistance of finance staff to new technologies and insufficient understanding of business logic by technical staff result in a disconnect between requirements analysis and system development. Traditional departmental barriers and assessment mechanisms fail to meet the needs of AI transformation, and cross-departmental collaboration is inefficient. The lack of a unified coordination mechanism at the project management level, business process reengineering encountered employee resistance, and change management became more difficult. Practice shows that organisational synergy has become a key bottleneck restricting the effectiveness of AI financial systems, requiring management to establish a more effective cross-departmental cooperation mechanism.

4.2.3. Cost-benefit Trade-offs

The payback period of AI financial systems is generally long, and companies face severe cost-benefit trade-offs when making project decisions. As shown in Figure 3, the system requires a large amount of capital investment in the preconstruction phase and does not begin to generate positive returns until the mature operation period. Market research shows that it takes an average of 18 months for a complete AI financial system to go into production from project inception, with a payback period of more than 36 months. There is significant uncertainty between system construction costs and expected returns, especially during economic downturns, when the willingness of enterprises to invest significantly declines.

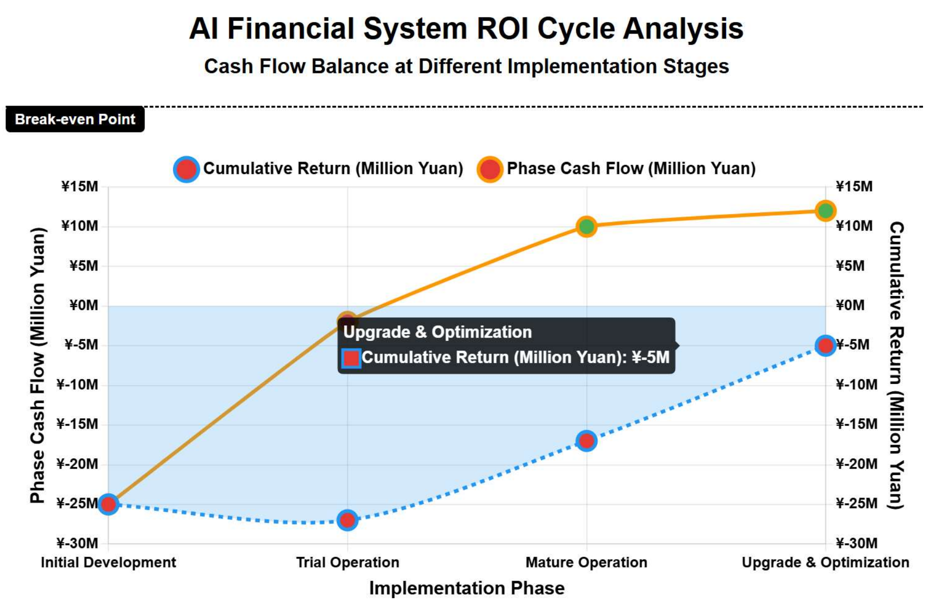

Figure 3. Analysis of the AI financial system payback cycle

Many enterprises suffer from a tight capital chain in the middle of the project, which leads to the shelving of system construction or the shrinking of functions. Accurate quantification of the intangible benefits brought by the AI system is still a problem plaguing enterprise decision-making, and a more scientific investment evaluation system needs to be established. Data shows that only about 40 per cent of enterprises are able to achieve the return on investment target within the expected time, and this percentage is even lower among SMEs.

5. Future Outlook and Suggestions

The future development of generative AI in the financial field presents three major trends: technological innovation, system improvement and research deepening. At the technology level, dynamic ethical engines and intelligent monitoring systems will become the focus of development, and the global market size is expected to exceed USD 200 billion in 2025, and AI systems will have stronger autonomous ethical assessment capabilities and data security protection capabilities. At the institutional level, regulators will establish a hierarchical regulatory framework, promote the development of industry self-regulatory standards, and strengthen international regulatory collaboration. At the research level, it will deepen the exploration of social impact, interdisciplinary integration and other dimensions, and promote the healthy interaction between theoretical and practical innovation. The synergistic development of these three dimensions will promote the evolution of AI financial systems in the direction of more intelligent, secure and responsible, and provide strong support for the construction of sustainable financial digital transformation.

References

[1] Joshi R, Pandey K, Kumari S. Generative AI: A Transformative Tool for Mitigating Risks for Financial Frauds[J]. Generative Artificial Intelligence in Finance: Large Language Models, Interfaces, and Industry Use Cases to Transform Accounting and Finance Processes, 2025: 125-147.

[2] Porwal P D. Ethics and Laws: Governing Generative AI's Role in Financial Systems[J]. Generative Artificial Intelligence in Finance: Large Language Models, Interfaces, and Industry Use Cases to Transform Accounting and Finance Processes, 2025: 283-298.

[3] Thuraisamy K S, Baig Z, Zeadally S. Generative Artificial Intelligence in Financial Services: Opportunities, Challenges, and Cyberthreats[J]. IT Professional, 2025, 27(2): 35-41.

[4] De Villiers C, Dimes R, Molinari M. How will AI text generation and processing impact sustainability reporting? Critical analysis, a conceptual framework and avenues for future research[J]. Sustainability Accounting, Management and Policy Journal, 2024, 15(1): 96-118.

[5] Nidhisree C, Paul A, Venunadh A, et al. Generative AI Under Scrutiny: Assessing the Risks and Challenges in Diverse Domains[C]//2024 IEEE 6th International Conference on Cybernetics, Cognition and Machine Learning Applications (ICCCMLA). IEEE, 2024: 243-248.

[6] Al-Hchemi L H. Evaluating Generative AI in enhancing banking services efficiency[C]//Economic Forum. Lutsk, Ukraine: Lutsk National Technical University, 2024, 14(4): 47-54.

[7] Wach K, Duong C D, Ejdys J, et al. The dark side of generative artificial intelligence: A critical analysis of controversies and risks of ChatGPT[J]. Entrepreneurial Business and Economics Review, 2023, 11(2): 7-30.

[8] Moharrak M, Mogaji E. Generative AI in banking: empirical insights on integration, challenges and opportunities in a regulated industry[J]. International Journal of Bank Marketing, 2025, 43(4): 871-896.

[9] Gowda P. Benefits and Risks of Generative AI in FinTech[J]. Journal of Scientific and Engineering Research, 2024, 11(5): 267-275.

[10] Gupta R, Pandey G, Pal S K. Automating government report generation: A generative ai approach for efficient data extraction, analysis, and visualization[J]. Digital Government: Research and Practice, 2025, 6(1): 1-10.

[11] Vuyyuru H K. Navigating ethical dilemmas in Generative AI Development: A framework for responsible innovation[J]. International Journal of Science and Research Archive, 2025, 14(1): 1416-1423.

[12] Beltran M A, Ruiz Mondragon M I, Han S H. Comparative analysis of generative ai risks in the public sector[C]//Proceedings of the 25th Annual International Conference on Digital Government Research. 2024: 610-617.

[13] Qi R. DecisionFlow for SMEs: A Lightweight Visual Framework for Multi-Task Joint Prediction and Anomaly Detection[J]. 2025.

[14] Wang Y. Construction of a Clinical Trial Data Anomaly Detection and Risk Warning System based on Knowledge Graph[C]//Forum on Research and Innovation Management. 2025, 3(6).

[15] Chen R. The application of data mining in data analysis[C]. Proceedings of the International Conference on Mathematics, Modeling, and Computer Science (MMCS2022), 2023, 12625: 473-478. SPIE.