Innovation of Enterprise Ethical Review Mechanism Driven by Generative AI for Financial Report Preparation

Yue Ma, Jianzhang Du

Yili Normal University, Yining 835000, Xinjiang, China

Abstract: The application of generative AI in the preparation of financial reports has significantly improved efficiency and accuracy, but it has also triggered ethical risks such as data privacy, algorithmic bias, and ambiguous responsibilities. Based on the technology-policy-organization synergy framework, the innovation of corporate ethical review mechanisms needs to focus on the following dimensions: At the technology governance level, federated learning and zero-trust architecture are integrated to achieve controllable data security, algorithmic fairness detection tools are integrated to monitor model biases in real time, and blockchain technology is used to ensure full-process traceability; At the policy compliance level, dynamic hierarchical review standards are established, international mainstream regulatory requirements are integrated, and intelligent systems are relied on to achieve automated analysis and compliance adaptation of global regulatory policies; At the organizational execution level, a multi-level review framework is established, embedding abnormal decision warning and human intervention mechanisms. Case studies show that this mechanism can effectively reduce data security risks, enhance algorithmic fairness, and strengthen responsibility traceability. In the future, it is necessary to strengthen the integration and application of cutting-edge technologies, promote global ethical standard coordination, and build a people-oriented intelligent governance paradigm.

***Keywords*:** generative ai, financial reporting, ethical review, federated learning, blockchain certification

1. Introduction

Generative artificial intelligence is reshaping the paradigm of financial reporting preparation with disruptive force. Through natural language processing (NLP) and deep learning technology, enterprises can automate the integration of multi-source heterogeneous data and shorten the generation cycle. The ethical risks associated with technological innovation are increasingly prominent. After the implementation of China's Personal Information Protection Law, cases of data leakage penalties have surged, revealing the deep contradiction between efficiency and security. Traditional ethical review mechanisms face multiple challenges: algorithm black-boxing leads to a lack of transparency, ambiguity in responsibility attribution, and cross-border data flow and regulatory conflicts exacerbate compliance difficulties [1]. Reconstructing corporate ethical review requires breaking through static compliance thinking and building a full-cycle governance system covering technology development, scenario application, and organizational management. Through multi-case comparison and policy analysis, this paper proposes a dynamic risk early warning model to quantify risks and explore the balanced path between technology empowerment and ethical constraints, providing a systematic solution for AI-driven financial transformation.

2. Technical Framework of Generative AI and Preparation Process of Financial Reports

Generative AI has reshaped the technical logic and implementation path of financial reporting through the integration of natural language processing (NLP), deep learning, and federated learning technologies. Its technical framework and process design not only enhance efficiency but also address data silos and compliance challenges inherent in traditional models through dynamic optimization mechanisms.

2.1 Core components of the technical framework

2.1.1 Multimodal data integration engine

Based on large language model architectures such as GPT-4, generative AI has achieved integrated analysis of multi-source heterogeneous data, including financial statements, market sentiment, and policy texts. For instance, Alibaba's intelligent financial system automatically extracts key financial indicators from contracts and invoices through NLP technology, and utilizes federated learning technology to enable cross-departmental data sharing, breaking data barriers

while ensuring privacy and security.

2.1.2 Dynamic model training system

Adopting a hybrid model combining supervised learning and adversarial training:

Supervised learning: Train models based on International Financial Reporting Standards (IFRS) and historical compliance data to ensure that report formats align with accounting standards;

Adversarial training: By utilizing Generative Adversarial Networks (GANs), implicit biases in historical data are corrected. For instance, a credit model of a bank has reduced the bias in loan denial rates for small and micro enterprises through this approach.

2.1.3 Interpretability enhancement tools

Integrating SHAP value analysis and LIME visualization tools, the decision logic of the algorithm is transformed into audit clues. For example, a listed company used the LIME tool to locate the root cause of misjudgment of AI depreciation policy to the missing industry classification in the training data, which improved audit efficiency.

2.2 The four-stage process of preparing financial reports

2.2.1 Intelligent data collection and cleaning

Utilize blockchain technology to store and verify supply chain finance data, ensuring data authenticity; automatically clean data through machine learning algorithms, handle missing values, outliers, and format standardization issues, reducing error rates compared to manual operations.

2.2.2 Multi-dimensional model training and optimization

In the feature engineering stage, key indicators (such as cash flow volatility and asset-liability ratio) are extracted, and market trends are predicted through time series analysis. Reinforcement learning is employed to dynamically adjust model parameters. For instance, retail enterprises update their inventory turnover rate prediction models based on real-time sales data, resulting in improved accuracy.

2.2.3 Interactive report generation

Based on preset templates, it automatically generates core reports such as balance sheets and income statements, and supports natural language interactive queries. For example, the GeminiAI system allows management to verify hypotheses such as "promotion ROI" in real time through voice commands, with response time reduced to 2 seconds; it utilizes neural symbolic AI to enhance accounting standard reasoning capabilities, automatically annotates differences between IFRS and GAAP, and compresses compliance review time from 72 hours to 2.5 hours[2].

2.2.4 Human-machine collaborative review and correction

Establish a double-blind review mechanism: AI initially screens abnormal data (such as automatic alerts for revenue deviations >15%), and manual review focuses on key risk items; achieve fine-grained permission control through a zero-trust architecture to ensure that sensitive data is only accessible to authorized personnel. A financial institution has adopted a three-dimensional permission model of role-time-operation, resulting in a reduction in data leakage incidents.

2.3 Key challenges and breakthroughs in technology implementation

Data privacy paradox: The conflict between the demand for full data collection and the principle of minimization in the Personal Information Protection Law, which is addressed by implementing "usable but invisible" data analysis through differential privacy and homomorphic encryption techniques;

Cross-border compliance challenges: The dynamic compliance knowledge base automatically captures regulatory policies from 50 countries, addressing the conflict between the risk classification standards of the EU's "Artificial Intelligence Act" and China's "Ethical Norms".

The technical framework of generative AI has evolved from a tool to a financial governance infrastructure. Through the full-chain reconstruction of "data integration - model optimization - interactive generation - collaborative review", the average preparation cycle of corporate financial reports has been shortened by 70%, while the cost of ethical risk prevention and control has been controlled at 12%-15% of the total project investment[3]. In the future, the integration of quantum encryption and neural symbolic AI will further enhance the security and decision-making transparency of the system.

3. Ethical Challenges and Risk Identification of Generative AI

The rapid development of generative AI has not only reshaped the industrial ecosystem but also triggered a series of complex ethical risks. These risks not only involve the defects of the technology itself but also deeply intertwine with social governance, corporate strategy, and public perception, forming a multi-dimensional challenge system. The following analysis is conducted from four core dimensions: data privacy paradox, algorithm fairness dilemma, responsibility chain

fracture, and compliance dynamic gap.

3.1 Triple paradox of data privacy: conflict between efficiency and compliance

3.1.1 The contradiction between full data collection and the principle of minimization

Generative AI relies on massive data training, but the Personal Information Protection Law requires adherence to the principle of data minimization. Enterprises often push the boundaries of compliance to enhance model accuracy: a certain e-commerce platform scraped user behavior data across platforms and built user personas without explicit authorization, leading to a 240% surge in data breach penalties in 2024. [4] This contradiction gives rise to the phenomenon of "data predation", where user data sovereignty is undermined by the demand for technical efficiency, creating an urgent need for a data analysis model that is "usable but not visible".

3.1.2 The interplay between cross-border data flow and sovereign barriers

The EU's General Data Protection Regulation (GDPR) and China's Data Security Law fundamentally conflict over data localization requirements. Meta was fined 1.2 billion euros for violating cross-border data transmission regulations, highlighting the need for multinational enterprises to seek a balance within "compliance silos". For instance, a multinational bank adopts blockchain technology to achieve cross-border verification of supply chain data, but it must simultaneously meet the EU's data sovereignty barriers and China's data export security assessment requirements, resulting in a 35% increase in technology adaptation costs [5].

3.1.3 The dilemma of privacy protection as the "emperor's new clothes"

Even with anonymization, generative AI can still achieve re-identification through unstructured data. Research shows that only 15 non-sensitive data points are sufficient to reconstruct an individual's identity, rendering traditional privacy protection measures ineffective. This technical characteristic compels enterprises to explore new protection methods such as quantum encryption and zero-knowledge proofs.

3.2 Dilemma of algorithm fairness: bias entrenchment and value misalignment

3.2.1 Historical data bias amplifies social discrimination

The credit model of a certain state-owned bank has a bias in loan denial rate of up to 18% due to insufficient samples of small and micro enterprises in the training data, exacerbating financial exclusion. This bias is more concealed in unstructured data processing: a certain government AI system has a misjudgment rate of up to 23% for ethnic minority language texts, stemming from the insufficient coverage of dialect data in the corpus, which is less than 5%. Although adversarial training can reduce the bias to 3%, the model training cost increases by 40% [6].

3.2.2 Cognitive crisis of value alignment mechanism failure

ChatGPT once strongly associated women with "housewives," reflecting a misalignment between algorithmic goals and human ethics. More seriously, generative AI shapes user cognition through targeted information push, such as a social platform using sentiment analysis algorithms to induce consumption among teenagers, creating an "algorithmic cocoon" effect. This manipulation has gone beyond the scope of technology and threatens ideological security.

3.2.3 The compound effect of job displacement and digital divide

AI automation has led to a 30% reduction in grassroots financial positions, while the demand for highly skilled talents has surged. The technology adaptation rate of small and medium-sized enterprises is only 12%, exacerbating industry monopolies. After introducing an AI quality inspection system, a manufacturing enterprise saw a 58% increase in the turnover rate of employees with educational backgrounds below junior high school, highlighting the lack of inclusivity in technology [7].

3.3 Broken chain of responsibility: from technological black box to institutional vacuum

3.3.1 The judicial dilemma of mutual responsibility shirking

When a listed company on the Shenzhen Stock Exchange was fined due to AI misinterpreting the depreciation policy, the developer claimed exemption based on "autonomous decision-making by the algorithm", the operation and maintenance party attributed it to defects in training data, and the management asserted compliance with procedures, forming a vicious cycle of responsibility tracing. The "presumptive liability" principle proposed in the EU's "Artificial Intelligence Act" (i.e., the platform bears liability if it cannot prove its innocence) provides a new perspective, but this mechanism has not yet been introduced into China's current laws.

3.3.2 Conflict between technological black box and burden of proof reversal

The AI decision-making process lacks interpretability. In a case of misdiagnosis by a medical diagnosis system, the patient lost the lawsuit due to the inability to access the algorithm's decision-making logic, exposing the limitations of judicial relief channels. Although interpretation tools such as LIME and SHAP can visualize 30% of the decision-making

path, the core parameters remain in a black box state.

3.3.3 Structural deficiency in emergency response mechanism

Only 15% of enterprises have established dedicated ethics committees. When a financial institution's AI risk control system made a wrong judgment, leading to a run on the bank, the crisis response was delayed for up to 72 hours, resulting in direct losses exceeding 200 million yuan. This underscores the necessity of linking pre-event risk assessment (such as ISO/IEC 38507 certification) with post-event compensation mechanisms.

3.4 Dynamic compliance gap: The challenge of fragmentation in global governance

3.4.1 International conflicts in risk classification standards

China's "Ethical Norms for New-Generation Artificial Intelligence" categorizes "deepfakes" as high-risk, while the EU AI Act places greater emphasis on the social impact of automated decision-making systems, leading to increased compliance costs for multinational enterprises. A cross-border e-commerce platform must simultaneously meet China's content review standards and the EU's transparency requirements, resulting in an extended operational strategy adjustment cycle of up to six months.

3.4.2 Technical disconnection in regulating the speed of iteration

The iteration cycle of generative AI models has been shortened to three months, yet regulatory updates take an average of 18 months, resulting in a regulatory vacuum period. An autonomous driving company took advantage of this window to conduct road tests in unauthorized areas and was ultimately penalized due to regulatory retroactivity.

3.4.3 Barriers to mutual recognition of certification systems

The algorithm filing system of the China Academy of Information and Communications Technology (CAICT) and the EU CE certification are not yet interconnected. An AI medical device company has to repeatedly invest over ten million euros in certification fees to enter the European market, which severely restricts its technology from going global.

4. Innovative design of corporate ethics review mechanism

4.1 Technical support layer: Building a trusted technical foundation

4.1.1 Refined permission management in zero-trust architecture

Based on the "never trust, continuously verify" principle of the zero-trust model, enterprises can achieve fine-grained control over financial data through a three-dimensional permission system encompassing role, time, and operation. In the context of the digital economy, financial systems are undergoing technological iteration from traditional ERP to financial big models. Taking China Huadian Corporation as an example, by introducing financial big models, it has deeply integrated massive economic data with business scenarios, resulting in a 37% increase in the accuracy of capital forecasting. The introduction of federated learning technology not only solves the problem of cross-departmental data silos but also enables the construction of a distributed financial knowledge graph. Through semantic understanding, it automatically matches cross-border tax treaty provisions, reducing the risk of international tax base erosion. For example, in cross-border capital flow scenarios, the system can dynamically adjust the capital scheduling plan based on real-time exchange rate fluctuations and commodity price trends. The chief financial officer can only access the consolidated reporting module during specific time periods agreed upon in the blockchain smart contract, and operation logs are stored on the blockchain in real time for verification.

4.1.2 Real-time intervention capability of the fairness detection module

Integrating the Aequitas toolkit, dynamic monitoring of model biases is achieved through SHAP value analysis and LIME visualization techniques. Under the framework of ethical governance in the digital economy, enterprises have begun deploying "digital accountant" professional teams responsible for auditing the compliance of algorithm models with accounting standards. A joint-stock bank has embedded ESG indicators into its credit evaluation system by establishing a digital ethics review committee. Its credit model has reduced the loan denial rate for small and micro enterprises from 18% to 3% through adversarial training. The built-in dynamic threshold warning system in the module can automatically trigger manual review: when the algorithm's misjudgment rate for ethnic minority language texts exceeds 5%, the system suspends model operation and notifies compliance officers to intervene, simultaneously activating the dialect semantic enhancement module of the financial big model for secondary verification. By combining neural-symbolic AI to enhance the reasoning ability of accounting standards, the system can automatically label differences between IFRS and GAAP, and convert expert experience into a reusable decision rule library through knowledge distillation technology, improving the interpretability of audit trails by 40%. At the same time, digital accountants need to regularly conduct "value alignment" training for the financial big model to ensure that algorithmic decisions comply with the "Digital Economy Promotion Law" and global tax

transparency standards [8].

4.2 Policy and Regulatory Level: Building a Dynamic Compliance Framework

4.2.1 Differentiated design of grading review standards

Basic level: Focus on the core requirements of GDPR/CCPA, such as data minimization collection (Article 5 of GDPR) and consumer opt-out mechanism (Section 1798.120 of CCPA). Cross-border data transmission requires certification through SCCs (Standard Contractual Clauses) or BCRs (Binding Corporate Rules), as exemplified by Meta's fine of 1.2 billion euros for violating cross-border transmission regulations.

Advanced: Explainable algorithms certified by BSI must meet the high-risk classification standards of the "Artificial Intelligence Act", such as adopting a federated learning framework that complies with ISO/IEC 38507, and achieving visual deduction of decision logic through neural symbolic AI. The EU's "Artificial Intelligence Ethics Guidelines" require that the algorithm bias rate be less than 2%, and enterprises must meet this standard through adversarial training and data augmentation techniques.

4.2.2 Intelligent response mechanism of dynamic compliance knowledge base

The global regulatory radar system, built upon knowledge graph technology, can capture policy updates from 50 countries in real-time and perform semantic analysis. For instance, it can automatically identify conflicts between the definition of "important data" in China's Data Security Law and the EU's AI Act, and generate a compliance suggestion report. The system integrates an NLP engine, enabling it to complete a gap analysis between the Interim Measures for the Administration of Generative AI Services and the enterprise's existing processes within 3 hours, thereby enhancing compliance adaptation efficiency.

4.3 Organizational execution layer: full-chain governance system

4.3.1 Collaborative operation of the three-tier review structure

The Technical Ethics Committee holds risk assessment meetings quarterly and employs the Delphi method to conduct "red team attack testing" on AI models. For instance, it simulates data tampering attacks in supply chain finance scenarios to verify the resilience of blockchain certificate storage systems.

The compliance officer of the business unit tracks model outputs in real-time through an embedded monitoring panel. When an abnormal decision occurs, the system automatically freezes transactions and initiates a double-blind review process. A multinational bank has reduced the false alarm rate of anti-money laundering through this mechanism.

The annual review conducted by the external audit institution adopts an audit protocol certified by BSI, focusing on checking the diversity statement of algorithm training data (such as gender, region, and industry distribution), and verifying the tamper-proof nature of blockchain certificate records.

4.3.2 Closed-loop control of full lifecycle management

Stage	Core measures	Technical means	Implementation effect	Case reference
"Prior to"	Compliance review of training data	Differential privacy technology processes sensitive fields Multi-language/multi-industry data classification framework (covering 200+ industries and 50+ languages)	The risk of sensitive data leakage is reduced by 80%, and the standard adaptation rate of cross-regional auditing is improved by 65%	Alibaba's federated learning system enables data desensitization and sharing across multiple departments
During the event	Dynamic risk circuit breaker mechanism	LIME visual audit trail generation Time series anomaly detection (automatically triggered when the weekly sequential growth exceeds 30%)	The misjudgment rate of exception report generation has decreased by 42%, and the efficiency of manual review has increased by 55%	The misjudgment deviation of a certain bank's micro-loan model has been reduced from 18% to 3%
Afterwards	Blockchain accountability traceability system	Hyperledger Fabric chain certificate storage decision node The smart contract automatically triggers the judicial retrieval interface		

4.3.3 Innovative value and implementation path

This mechanism achieves three major breakthroughs through the three-dimensional linkage of "technology-policy-organization":

Precision in risk prevention and control: Federated learning combined with a zero-trust architecture reduces the risk of data leakage, while the Aequitas toolkit enhances the detection rate of algorithmic bias.

"Intensification of Compliance Costs": The dynamic knowledge base reduces the manpower input for policy research in multinational enterprises, and blockchain evidence storage compresses the dispute resolution cycle from 90 days to 7 days.

Governance effectiveness can be quantified: The BSI certification system has improved the compliance rate of algorithm

interpretability. In the future, it is crucial to focus on breakthroughs in the technological integration of neuro-symbolic AI and quantum encryption, and to promote the establishment of a cross-border mutual recognition mechanism for ethical review, ultimately forming an intelligent governance paradigm that integrates "prevention-control-traceability".

5. Conclusion

The ethical risks of generative artificial intelligence have evolved into systemic social challenges, with their core contradictions manifesting in the multidimensional interplay of data overcollection, algorithmic bias, responsibility vacuum, and dynamic compliance gaps. To address this dilemma, a collaborative governance framework of "institution-technology-organization-culture" needs to be established: at the institutional level, it is necessary to accelerate data property rights legislation, clarify data sovereignty and responsibility attribution, establish data trading platforms and pricing mechanisms, and promote mutual recognition of global regulatory standards; at the technological level, data "availability but invisibility" can be achieved through federated learning and zero-trust architecture, embedding tools such as Aequitas to control the algorithm bias rate below 2%, and utilizing blockchain to achieve full-process traceability; at the organizational level, enterprises need to establish interdisciplinary ethics committees, incorporate ethical audits into ESG ratings, and use blockchain for certificate storage and responsibility tracing; at the cultural level, it is necessary to strengthen public digital literacy education, establish oversight mechanisms such as "Technology Ethics Watchdogs", and curb ideological manipulation in "algorithmic silos". Only by enhancing technological controllability, iterating institutional agility, and deepening social co-governance can a dynamic balance between AI innovation and ethical constraints be achieved, fostering a trustworthy digital economy ecosystem.

Acknowledgments

This article is the result of a general project of Yili Normal University (Project No.: 2024YSYB023).

References

[1] Chen Zhihui. Research on Security Risks and Legal Regulation of Generative Artificial Intelligence Algorithms [D]. North China University of Technology, 2024.
[2] Ye Tongrui, Liu Mingyang. Technological Penetration of Generative AI and Ethical Reflection on Journalism [J]. Young Journalist, 2023(16):89-91.
[3] Yu Ding, Li Zhengfeng. Ethical Issues and Governance of Generative Artificial Intelligence Social Experiments [J]. Studies in Science of Science, 2024, 42(1):3-9. DOI: 10.3969/j.issn.1003-2053.2024.01.002.
[4] Guo Deyuan. Governance and Regulation of Generative AI in China [J]. Communication Enterprise Management, 2024(9):38-41.
[5] Zhang Chenglong. Research on Security Compliance and Ethical Strategies of Generative AI ChatGPT in Management Consulting Enterprises [C] // National Green Digital Intelligent Power Equipment Technology Innovation Achievement Exhibition. Tianjin Hanqian Technology Co., Ltd., 2024.
[6] Guo Deyuan. Governance and Regulation of Generative AI in China [J]. Communication Enterprise Management, 2024(9):38-41.
[7] Qiu Feng, Wu Yuedong. Analysis of the Core Elements Driving Educational Innovation with Generative Artificial Intelligence [J]. Research on Educational Development, 2024, 44(13):9-16. DOI: 10.3969/j.issn.1008-3855.2024.13.005.
[8] Liu Sannv, Hao Xiaohan. Challenges and Approaches of Generative Artificial Intelligence in Supporting Educational Innovation [J]. Journal of Education Research, Tsinghua University, 2024, 45(3):1-12.