Available on: https://jefrr.ep-journals.org/ E-: 3051-3650 P-ISSN: 3051-3642

Journal of Economic Finance Research and Review

Volume: 02 Issue: 04 April 2026

Artificial Intelligence + Financial Risk Control: Application and Development Analysis of Credit Default Prediction Models Based on Machine Learning and Deep Learning

Deng Ke

School of Finance, Central University of Finance and Economics

Published Date: April 27, 2026 **Article DOI: 10.65150/EP-jefrr/V2E4/2026-06**

ABSTRACT: Driven by the "AI+" national strategy, artificial intelligence technologies—represented by machine learning, deep learning, and large language models (LLMs)—are revolutionizing the core paradigm of financial risk management with unprecedented depth. This paper focuses on the critical task of credit default prediction, systematically tracing the evolution of credit risk assessment techniques from traditional statistical learning to integrated machine learning and deep learning, while examining their theoretical boundaries and trade-offs in prediction accuracy, model interpretability, and complex data representation capabilities. Through case studies of WeBank, MYbank, and Du Xiaoman Finance, the study demonstrates three cutting-edge paradigms: privacy computing that breaks down data compliance silos, high-dimensional behavioral data-driven microcredit optimization, and LLM-powered unstructured risk reasoning. The research reveals that AI has evolved beyond being merely an efficiency-enhancing tool for risk control into a critical infrastructure that expands the reach of inclusive financial services through alternative data mining. However, practical implementation faces challenges including data compliance dilemmas, conflicts between model "black-box" characteristics and regulatory requirements, and algorithmic bias-induced heterogeneity effects. Finally, the paper envisions a next-generation intelligent risk management ecosystem—more open, sophisticated, and responsible—driven by the synergy of privacy computing, graph neural networks, and large language models. This study provides a systematic analytical framework for understanding the technical logic, application value, and developmental trajectory of intelligent risk management.

KEYWORDS: artificial intelligence; credit default prediction; machine learning; deep learning; large language models; financial risk control

1. INTRODUCTION

Currently, artificial intelligence (AI) technologies represented by machine learning, deep learning, and large language models (LLMs)are permeating all sectors of the economy and society with unprecedented breadth and depth. As a quintessential data-intensive industry, finance has emerged as the primary testing ground and value hub for AI applications due to its inherent demands for risk quantification and automated decision-making. Throughout this process, credit risk management remains central to maintaining financial system stability. Credit defaults not only directly erode the asset quality of financial institutions but may also trigger systemic macro-financial turbulence through their ripple effects. Therefore, continuously enhancing the accuracy and foresight of credit default prediction remains an enduring core challenge for the financial sector.

As China's financial services enter a new phase of "grassroots penetration and inclusive access," credit customer segments are becoming increasingly fragmented and diversified. Traditional risk control models heavily rely on static financial metrics, historical credit records, and expert judgment. While conventional statistical frameworks like logistic regression and their derived scoring card systems demonstrated robust stability and interpretability when processing structured, low-dimensional financial data, they now face theoretical limitations in predictive accuracy when dealing with the massive, high-dimensional, and multi-modal "digital footprints" and alternative data emerging in the digital economy era. Their linear or generalized linear modeling assumptions struggle to capture complex nonlinear interactions, resulting in inadequate risk identification capabilities for emerging customer segments and intricate scenarios.

Meanwhile, the external environment poses dual challenges to the development of intelligent risk control. Technologically, the increasing complexity of models and their "algorithmic black box" nature create inherent conflicts with the financial industry's stringent regulatory compliance and consumer protection requirements (such as providing loan rejection justifications), resulting in the "precision-explainability trade-off." Socially and regulatory-wise, the full implementation of the Data Security Law and Personal Information Protection Law has intensified tensions between data's value as a critical production factor and the need to protect personal privacy and ensure data security. Cross-institutional and cross-domain data integration and collaborative modeling face severe "data silo" challenges.

Against this backdrop, exploring how to leverage cutting-edge AI technologies to address three core challenges—data compliance, dimensional explosion, and cognitive reasoning—has become an intrinsic need and inevitable path for driving high-quality development in the financial sector and expanding the boundaries of inclusive financial services. This study emerges at the intersection of technological transformation, business innovation, and regulatory reshaping. Under the strategic framework of "AI+", the research aims to systematically analyze the fundamental technical logic, typical application paradigms, practical value, and future challenges of AI technology in credit default prediction. Specifically, the study unfolds through a logical progression of "technological evolution—case deconstruction—value analysis—challenge outlook."

First, this paper systematically traces the evolution of credit default prediction technologies from traditional statistical models to ensemble machine learning and deep learning, conducting in-depth comparative analyses of theoretical advantages, practical challenges, and applicability boundaries across various models. Second, through deconstructing three technically representative cases—WeBank, MYbank, and Du Xiaoman Finance—we thoroughly examine the implementation logic and value creation models of privacy computing, behavioral data mining, and large language models in risk control. Building on these analyses, we identify and demonstrate the core value of AI-powered credit risk control in three dimensions: enhancing risk identification accuracy, expanding financial service accessibility, and strengthening dynamic monitoring capabilities. Subsequently, the study addresses critical challenges in current practices regarding data compliance, model explainability, and algorithmic fairness, while exploring integration directions for cutting-edge technologies including privacy computing, graph neural networks, and large language models. Finally, the paper summarizes key conclusions and provides targeted recommendations for financial institutions and industry ecosystems.

2. EVOLUTION OF TECHNICAL FRAMEWORKS FOR CREDIT DEFAULT PREDICTION: FROM STATISTICAL LEARNING TO COGNITIVE INTELLIGENCE

Credit default prediction fundamentally constitutes a binary classification problem, aiming to accurately assess the probability of default occurrence within specific future periods based on multi-dimensional characteristics of borrowers (individuals or enterprises). Technological advancements have consistently focused on enhancing predictive performance through more comprehensive and intelligent data utilization. This section systematically reviews the technical evolution from traditional statistical learning to modern ensemble machine learning, and further to deep learning and hybrid architectures, while providing in-depth analysis of the underlying theoretical logic and trade-offs.

2.1 Traditional Statistical Models: The Quantitative Foundation and Theoretical Boundaries of Credit Assessment

Prior to the widespread adoption of AI technology, credit risk assessment relied extensively on traditional statistical models such as logistic regression, discriminant analysis, and survival analysis. Among these, logistic regression emerged as the cornerstone for constructing credit scoring cards due to its simple model structure, parameters with clear statistical significance (e.g., the Odds Ratio), probabilistic interpretability, and relative robustness against multicollinearity among features.

Extensive early research has validated the effectiveness of traditional models under specific conditions. For instance, Yang Pengbo et al. (2009) developed a credit default prediction model for listed companies using Logistic regression, achieving high classification accuracy on test sets by screening core financial indicators such as return on assets and current liability ratios. Guo Xinwei's (2012) study on small and medium-sized enterprises also demonstrated that Logistic regression models could achieve over 90% prediction accuracy in small-sample tests after rigorous financial indicator screening. The advantage of these models lies in their "white-box" characteristics: coefficients and their significance can be rigorously tested, with the entire decision-making process offering global traceability. This perfectly aligns with regulatory requirements for model risk management transparency under strict frameworks like the Basel Accord.

However, the theoretical assumptions of traditional statistical models (such as linearity and feature independence) also define their application boundaries. With the expansion of financial service customer bases and the explosion of data dimensions, their limitations have become increasingly apparent: First, they struggle to effectively model complex high-order nonlinear interaction effects; Second, feature engineering heavily relies on expert experience, making it difficult to automatically extract meaningful features from massive alternative datasets; Third, they exhibit virtually zero capability in processing unstructured data (e.g., text and images). Consequently, when dealing with the massive, sparse, and high-dimensional behavioral data generated by internet ecosystems, traditional models are prone to underfitting, resulting in significant limitations in predictive accuracy.

2.2 Integrated Machine Learning: Overcoming Bottlenecks in Nonlinear and High-Dimensional Features

To overcome the theoretical limitations of classical statistical methods, machine learning models represented by ensemble learning have gradually become mainstream in both academic research and industrial applications. The core concept of ensemble learning involves constructing and combining multiple "weak learners" to form a "strong learner." Among these, gradient-boosting decision tree algorithms based on the boosting framework—such as XGBoost, LightGBM, and CatBoost—have become the de facto standard for processing structured credit tabular data due to their outstanding predictive performance, high computational efficiency, and robustness to missing values and outliers.

The theoretical advantage of this model lies in its non-parametric nature, which eliminates the need for stringent prior assumptions about data distributions and enables automatic learning of complex nonlinear decision boundaries and high-order feature interactions through tree structure combinations. Its practical value has been extensively validated through empirical studies. For instance, research by Carbo and Alonso (2021) demonstrated that in real-world credit prediction scenarios, XGBoost outperforms traditional logistic regression models significantly in predictive accuracy; this improvement translates into substantial savings in risk-weighted asset requirements for banks when applied to regulatory capital calculations[Error! Unknown switch argument.]. Further research by Li Yinqian (2025) showed that when incorporating multimodal alternative data such as mobile device behavior and multi-source credit inquiries, the LightGBM model achieved significantly higher AUC and KS values compared to models relying solely on traditional credit data, marking a shift in risk management research focus from "pure financial metric analysis" to enhancing discrimination capabilities for gray-market clients through digital footprint mining.

However, the advantages of integrated models come with challenges in interpretability. While post-hoc attribution tools like SHAP and LIME can provide local or global explanations for feature importance, such interpretations rely on approximate computations, lack rigorous causal logic support, and incur high computational costs. When facing regulatory inquiries or needing to provide consumers with intuitive loan rejection reasons, their explanatory power remains inferior to that of logistic regression.

2.3 Deep Learning and Hybrid Architectures: Decoding Complex Temporal-Spatial Correlations and Deep Semantic Representations

In recent years, as financial data has become increasingly diverse, time-series, and unstructured, deep learning has emerged as a key focus in risk control research due to its powerful end-to-end representation learning capabilities. Deep learning models such as deep neural networks, convolutional neural networks, and recurrent neural networks (particularly LSTM) can automatically extract abstract features from raw data through multi-layer nonlinear transformations, progressively capturing features from lower to higher dimensions. This approach significantly reduces reliance on labor-intensive manual feature engineering processes.

Among these, recurrent neural networks and their variants demonstrate exceptional performance in processing time-series data commonly encountered in credit scenarios—such as repayment records, transaction sequences, and APP activity trends—effectively capturing long-term dependency patterns in historical behaviors to enable dynamic risk assessment. Current cutting-edge research increasingly focuses on constructing hybrid deep learning architectures that leverage the strengths of diverse models. For instance, Chen Wei et al. (2025), in their study on green credit default risk prediction for listed companies, proposed an AE-MA-CNN-LSTM hybrid model[Error! Unknown switch argument.]. This model employs autoencoders for dimensionality reduction and a multi-head attention mechanism to capture global correlations, while integrating CNN and LSTM to extract spatiotemporal features, thereby addressing the challenge of complex nonlinear prediction tasks.

Deep learning demonstrates unparalleled advantages in processing massive, complex, and high-noise alternative data (e.g., images, texts, and complex sequences). A comparative study by Rakshith Bhandary and Bidyut Kumar Ghosh (2025) confirmed that deep neural networks outperform traditional methods across comprehensive metrics in ultra-large-scale credit card default prediction, even surpassing ensemble models in specific scenarios. However, their pronounced "black-box" characteristics—where model decision-making processes lack intuitive understanding—pose significant compliance and ethical challenges in credit decision-making contexts requiring high transparency and strong interpretability.

2.4 Comparison of Technical Approaches and Integration Trends

In summary, statistical learning, ensemble machine learning, and deep learning collectively form a complementary technical spectrum with distinct focuses in credit default prediction.

- The core strengths of statistical learning methods (e.g., logistic regression) lie in their interpretability and regulatory compliance-friendly nature, featuring transparent models and well-defined parameters, making them a stable choice for small-sample and heavily regulated scenarios. However, their predictive accuracy is subject to a ceiling effect.

- Integrated machine learning methods (e.g., XGBoost and LightGBM) have achieved an excellent balance between prediction accuracy and efficiency when processing structured tabular data, making them the undisputed mainstream in the industry. However, their interpretability is a 'post hoc approximation,' and their capability to handle unstructured data remains limited.
- Deep learning (e.g., DNN, LSTM) offers unique advantages in representing complex, unstructured, and multimodal data, representing the direction of cognitive intelligence. However, it faces challenges such as poor interpretability and a high risk of overfitting with small sample sizes.

The prevailing academic and industrial consensus holds that these three domains are not merely substitutable but demonstrate significant convergence trends. Future research frontiers lie in constructing hybrid intelligent architectures: leveraging deep learning (or large language models) with their robust feature extraction and semantic understanding capabilities to process complex raw data such as text, graphs, and time series, transforming them into structured risk features or semantic vectors. These high-quality features are then fed into ensemble learning models or lightweight frameworks incorporating explainability modules for final prediction and decision-making. This collaborative paradigm—where deep learning/LLMs handle feature engineering and complex cognition, while traditional models focus on precise prediction and interpretive outputs—is expected to strike an optimal balance between accuracy and explainability.

3. CASE STUDY ANALYSIS OF ARTIFICIAL INTELLIGENCE EMPOWERED CREDIT RISK CONTROL

The value of technology must be validated in real-world business scenarios. To thoroughly elucidate how AI is implemented across different business models, this section examines case studies of three institutions: WeBank, MYbank, and Du Xiaoman Finance. These institutions exemplify innovative practices at three progressive levels: leveraging AI to break down data silos, revolutionizing credit assessment paradigms, and enhancing risk perception dimensions.

3.1 WeBank's "Weilidai": A Privacy Computing Paradigm Based on "AI Downstreaming + Cryptographic Framework"

The core challenge faced by WeBank's "Weilidai" is serving a massive internet "long-tail customer base," where users often lack robust financial data such as central bank credit reports. To accurately assess their credit risks, it is essential to integrate external internet behavioral data, which directly conflicts with data compliance requirements under the Data Security Law and Personal Information Protection Law.

WeBank's innovative practice lies in transforming AI from an "upper-layer application algorithm" to a "bottom-layer computing engine," deeply integrating cryptographic technologies to establish a privacy-preserving risk control system centered on federated learning and multi-party secure computation. The theoretical breakthrough of this paradigm lies in revolutionizing the traditional "data remains static while models evolve" collaboration model. Within the federated learning framework, participating entities (such as banks and internet platforms) retain their data locally without physical transmission or centralized processing. Institutions exchange only intermediate training parameters (e.g., gradients and weights) through encryption protocols like homomorphic encryption and differential privacy. Central servers aggregate these encrypted parameters to update global models, which are then deployed. Through iterative cycles, this approach ultimately yields a joint risk control model with performance approaching centralized training systems.

Zhou Chuanxin et al. (2021) demonstrated that federated learning fundamentally represents a deep integration of AI optimization algorithms with modern cryptography. The success of WeBank exemplifies how the "AI decentralization + cryptographic framework" approach enables "data accessibility without visibility, and value creation without data manipulation." While strictly adhering to compliance requirements, this methodology effectively resolves data silos and achieves precise risk pricing for grassroots customer segments. Notably, privacy computing has evolved from a defensive "compliance cost" to a strategic "empowerment cornerstone" driving business growth.

3.2 MYbank's "310" Model: Credit Reconstruction for Small and Micro Enterprises Based on Behavioral Data

MYBank serves small and micro enterprises lacking standardized financial statements and collateral, where traditional risk control models prove largely ineffective. Its "310" model (3-minute loan application, 1-second disbursement, zero manual intervention) fundamentally replaces traditional financial data with massive, authentic behavioral data, leveraging machine learning models to systematically reshape the credit assessment of these enterprises.

The core of its risk control logic lies in transforming credit evaluation sources. Leveraging its ecosystem advantages, MYbank converts real-time operational data generated by small and micro enterprises—such as online store transaction records, warehousing logistics information, supply chain data, and customer reviews—into credit assets. These high-dimensional, dynamic datasets provide far more accurate insights into a company's true operational status and cash flow health compared to static reports. By employing gradient boosting tree models like LightGBM, MYbank efficiently processes these high-dimensional sparse features, automatically identifying complex nonlinear relationships and interaction effects to precisely quantify corporate credit risks.

This model has achieved dual breakthroughs: In terms of inclusiveness, it extends financial services to tens of millions of small and micro entities—often referred to as "credit white users" —that were previously inaccessible to traditional finance. Regarding risk control, it leverages real-time behavioral data to deliver superior asset quality compared to conventional models, demonstrating that "inclusive finance does not inherently entail high risks." This marks a shift in risk management from historical outcome-based assessments to real-time behavioral prediction systems.

3.3 The "Xuanyuan" Financial Large Model for the 3.3 Degree Minor Full Moon: LLM-Driven Unstructured Risk Perception

Traditional machine learning excels at processing structured data, yet over 70% of valuable information in financial risks resides in unstructured texts—such as corporate financial notes, legal documents, public sentiment reports, and due diligence materials. Conventional NLP technologies have limited capacity to handle such information. Leveraging its billion-scale Chinese financial large model "Xuanyuan," Duxiaoman has pioneered a novel credit risk management paradigm grounded in cognitive reasoning.

The core capabilities of large language models (LLMs)lie in deep semantic understanding, logical reasoning, and zero-shot/low-shot learning. In complex corporate credit assessments or due diligence processes, LLMs can serve as "super AI analysts" that swiftly process hundreds of pages of documents to identify critical insights such as related-party guarantees, risks associated with actual controllers, and policy impacts, while performing cross-document logical deductions. For instance, analyzing a news report about a supplier suspending production due to environmental compliance requirements can reveal potential supply chain risks for downstream borrowing enterprises.

In practice, to balance the capabilities of large models with the rigor and interpretability required for financial decision-making, the industry widely adopts a hybrid bilateral architecture combining "LLM + traditional machine learning". Within this framework, large models do not directly make credit decisions but serve as powerful "feature engineering engines" and "information extractors". They transform unstructured text into structured risk labels, sentiment vectors, or knowledge graph relationships, then feed these high-quality features into models with stronger interpretability (e.g., LightGBM) for final decision-making. This architecture not only significantly expands the information input boundaries for risk control but also ensures decision process stability and accountability through traditional models, effectively mitigating the "hallucination" risks inherent in large models.

4. CORE VALUE ANALYSIS OF ARTIFICIAL INTELLIGENCE EMPOWERING CREDIT RISK CONTROL

Through technological evolution and case studies, AI's empowerment of credit risk management has progressed from merely enhancing operational efficiency as a tool to becoming a core infrastructure that drives business model transformation and expands financial service boundaries. Its value is primarily reflected in three key dimensions.

4.1 Stepwise Improvement in Risk Identification Accuracy and Operational Efficiency

AI models can automatically extract complex nonlinear patterns from massive high-dimensional datasets, significantly enhancing risk differentiation capabilities. Empirical studies demonstrate that advanced ensemble learning models outperform traditional logistic regression by 5-15 percentage points in AUC values, with notable improvements in KS values. This indicates the models can more accurately distinguish between high-performing and high-risk customers, reducing false rejections (improving approval rates) and customer attrition (lowering bad debt rates). Such precision gains directly translate into operational efficiency breakthroughs. Leading institutions have achieved automated approval rates exceeding 90%, compressing traditional processes from hours to seconds, substantially lowering operational costs while freeing human resources to focus on optimizing complex model strategies and analyzing high-risk cases.

4.2 A Fundamental Expansion of the Inclusiveness of Financial Services

The most profound contribution of AI lies in resolving the "information asymmetry" dilemma in traditional risk management. By utilizing "alternative data," it establishes creditworthiness for "white households" lacking credit records. For individuals, digital footprints (including payments, consumption patterns, and social interactions) serve as credit benchmarks; for small and micro enterprises, operational data (such as transaction records and supply chain information) form the foundation of credit evaluation. As demonstrated by MYbank's case, this approach enables financial services to reach hundreds of millions of long-tail customers outside traditional financial systems. Simultaneously, precise risk pricing capabilities make differentiated, risk-controlled inclusive financial services feasible, breaking the outdated paradox that "inclusiveness equates to high risk and high costs," thereby achieving commercial sustainability.

4.3 Systematic Enhancement of Dynamic Risk Monitoring and Early Warning Capabilities

AI drives risk management evolution from static pre-loan assessments to dynamic, end-to-end continuous oversight. Leveraging stream computing and real-time feature engines, the model enables ongoing monitoring of post-loan client behavior, swiftly detecting early risk signals like abrupt transaction patterns or behavioral anomalies. This transformation shifts risk management from "post-event collection" to "in-process intervention" and even "pre-event alerts." The intelligent collection system formulates differentiated strategies based on repayment probability predictions, significantly boosting recovery efficiency. Such advancements transform risk management from reactive responses to proactive defense and precision operations, substantially enhancing the resilience of entire asset portfolios against risks.

5. REAL-WORLD CHALLENGES AND FUTURE TECHNOLOGICAL PROSPECTS

Despite its promising prospects, the comprehensive application of intelligent risk control still faces complex challenges across three dimensions—technology, ethics, and regulation—while its future development clearly demonstrates a trend toward integrated evolution.

5.1 Core Challenges Currently Facing

First, the balance between data compliance and value utilization remains challenging. While privacy-preserving computing offers potential solutions, it still faces hurdles in performance degradation, communication overhead, and engineering complexity, requiring breakthroughs for large-scale commercial implementation. Second, the inherent conflict between model prediction accuracy and interpretability constitutes a core dilemma in financial scenarios. The "black box" nature of deep learning and large models creates tension with regulatory demands for decision transparency, fairness reviews, and consumer rights protection. Tools like SHAP remain inadequate in providing causal explanations. Finally, algorithmic biases and equity issues in financial resource allocation cannot be overlooked. Models trained with historically biased data may perpetuate or amplify discrimination. Research by She Kaiwen et al. (2024) warns that digital social credit platforms, while reducing overall default rates, may generate heterogeneous effects that exclude groups lacking digital footprints from benefits, potentially increasing financing costs and exacerbating the "digital divide".

5.2 Future Directions in Cutting-edge Technology

To address these challenges, next-generation intelligent risk control technologies are evolving toward integration and collaboration.

1. The scaling and standardization of privacy computing: Technologies such as federated learning, multi-party secure computation, and trusted execution environments will achieve deep integration, evolving toward more efficient and standardized interoperability protocols to establish cross-institutional risk joint prevention and control networks.
2. The profound applications of Graph Neural Networks (GNN): GNN naturally models the complex network relationships among customers, enterprises, and transactions, significantly enhancing the ability to identify and penetrate gang-related fraud, associated guarantee risks, and systemic risk transmission pathways—achieving a cognitive leap from "individual risk" to "network risk."
3. The vertical domain specialization of large language models (LLMs): In the financial sector, these large models will continuously enhance their accuracy and reliability in scenarios such as risk report analysis, compliance review, and public opinion risk insights through technologies like retrieval-enhanced generation, knowledge graph integration, and fine-tuning with high-quality financial corpora. They will also achieve deeper integration with traditional models within a "dual-architecture" framework.

5.3 Reconstruction of Future Ecosystems

The evolution of technology will ultimately drive the restructuring of businesses and their ecosystems. Risk management will transition from an isolated approval process to an intelligent core that is seamlessly integrated throughout the customer's entire lifecycle journey. Under the open banking paradigm, leveraging privacy-preserving computing and standardized APIs, financial institutions, technology companies, and data sources will collaborate to establish an open, collaborative, and co-governed intelligent risk management ecosystem. Within this ecosystem, models, features, and insights can be securely shared and exchanged while ensuring security and privacy, collectively enhancing the financial system's risk identification and mitigation capabilities and fostering a more robust, inclusive, and responsible digital finance ecosystem.

6. CONCLUSION AND RECOMMENDATIONS

6.1 Research Conclusions

Under the national strategic framework of "Artificial Intelligence+", this study systematically analyzes the technological evolution, application paradigms, and future trends of AI in credit default prediction, yielding the following key conclusions:

First, technological evolution has transitioned from the "algorithmic precision race" to a new phase of "multimodal fusion and cognitive reasoning." From the interpretability foundation of statistical learning, to nonlinear breakthroughs in machine learning, and then to representation learning in deep learning, the ultimate goal is advancing toward semantic cognition in LLMs.The core trajectory involves continuously enhancing the ability to process and comprehend complex, multi-dimensional risk information.

Secondly, the key to successful implementation lies in "scenario adaptation" and "architectural integration." WeBank's privacy computing paradigm addresses challenges in compliant data flow; MYbank's behavioral data model redefines credit assessment logic for small businesses; while Duxiaoman's LLM hybrid architecture overcomes cognitive limitations in processing unstructured information. These three cases demonstrate that there are no universal "silver bullet" models—only solutions tailored to specific scenarios and data conditions. Hybrid architectures represent the mainstream approach to balancing performance and constraints.

Third, the value of AI has evolved from a defensive tool for "cost reduction and efficiency enhancement" to an offensive "business model innovation engine" and "inclusive financial infrastructure." By unlocking the value of alternative data, it fundamentally expands the potential boundaries of financial services.

6.2 Countermeasures and Recommendations

Looking ahead, to ensure the steady and sustainable development of intelligent risk management, the following recommendations are proposed:

1. Accelerate the deployment of privacy computing infrastructure: Financial institutions should regard privacy computing as a core strategic capability, actively invest in R&D and infrastructure development, and promote the establishment of industry-standard interfaces to facilitate secure data element circulation while ensuring compliance with security requirements.
2. Building an "Explainable AI" governance framework: Model explainability and fairness auditing must be elevated to strategic priorities on par with prediction accuracy. During model design, prioritize explainable architectures (e.g., hybrid models). Post-deployment, continuously employ tools like SHAP and LIME for monitoring and auditing, and establish algorithm impact assessment mechanisms to ensure transparent and equitable decision-making.
3. Pragmatically advance the vertical financial applications of large language models: Avoid blindly pursuing parameter scale; instead, focus on enhancing the accuracy, reliability, and controllability of LLMs in professional risk management scenarios through high-quality domain-specific corpus fine-tuning, retrieval-enhanced generation, and integration with knowledge graphs and traditional models (bilateral architectures), thereby transforming them into practical productivity tools.

6.3 Limitations and Future Perspectives

This study was constrained by commercial confidentiality and data accessibility, preventing quantitative empirical analysis of specific parameters in institutional models, feature engineering details, and micro-level transmission mechanisms of algorithmic biases. Future research could focus on three key directions: First, developing a quantitative framework to empirically examine Pareto optimal solutions for the "precision-explainability trade-off" under varying regulatory intensities; Second, exploring the integration of graph neural networks with large language models for mitigating systemic financial risk contagion; Third, conducting more detailed empirical studies to evaluate the heterogeneous impacts of intelligent risk control systems across socioeconomic groups, thereby informing the development of more inclusive digital financial policies.

REFERENCES

1) Chen Wei, Zhou Quanshi, Chen Zhensong, et al. Research on Green Credit Default Risk Prediction for Listed Companies Based on Hybrid Deep Learning Models [J]. Management Review, 2025,37(09):27-41.
2) Li Yinqian. Research on Credit Default Prediction Based on the LightGBM Model and Multimodal Data [J]. Financial Customer, 2025, (11):14-16.
3) She Kaiwen, Shen Yu, Zhao Shaoyang. The Impact of Big Data on Bank Credit Behavior: Evidence from the Digital Social Credit Platform [J]. Economic Research, 2024,59(03):147–165.
4) Yang Pengbo, Zhang Chenghu, Zhang Xiang. Research on a Credit Default Probability Prediction Model for Listed Companies Based on Logistic Regression Analysis [J]. Economic Geography, 2009, (02):144–148.
5) Guo Xinwei. Research on Credit Risk Measurement for Small and Medium-sized Enterprises in China [D]. Tianjin: Nankai University, 2012.

6) Zhou Chuanxin, Sun Yi, Wang Degang, et al. Review of Federal Learning Research [J]. Journal of Network and Information Security, 2021,7(05):77-92.
7) Bhandary, R., & Ghosh, B. K. Credit Card Default Prediction: An Empirical Analysis on Predictive Performance Using Statistical and Machine Learning Methods[J]. Journal of Risk and Financial Management, 2025, 18(1), 23.
8) Alonso, A., & Carbó, J. M. Understanding the performance of machine learning models to predict credit default: a novel approach for supervisory evaluation[J]. Banco de España Working Paper, 2021, No. 2105.

Corresponding Author: Deng Ke