Vol. 13 (2026): Proceedings of the 3rd International Conference on Business, Management and Sustainability (ICBMS 2026)
E-ISSN: 3105-0522 P-ISSN：3105-0514
DOI: https://doi.org/10.70267/icbms.2601.101108

Research on the Development and Optimization of Financial Investment in the Context of Big Data

Shang Chen*

Accounting and Financial Investment, Xiamen University

**Corresponding author: Shang Chen*

Abstract

With the deep integration of big data and artificial intelligence technologies, the field of financial investment is undergoing a profound paradigm shift. This study takes the practical performance of the domestic large model DeepSeek in the "Alpha Arena" AI Trader Competition as an example, and through literature review and empirical analysis, systematically explores the technical logic and optimization pathways of intelligent financial investment. The study finds that the DeepSeek model, by integrating multi-source heterogeneous data, employing nonlinear deep learning algorithms, and implementing dynamic risk control mechanisms, has achieved a fundamental transformation of investment strategies from experience-driven to data-intelligence-driven. Based on this, the present study proposes a five-dimensional optimization framework encompassing data factor circulation, enhancement of algorithm transparency, transformation of market structure, reconstruction of regulatory adaptability, and full-cycle risk management, aiming to provide theoretical support and practical guidance for the digital transformation of financial institutions.

Keywords

big data, financial investment, DeepSeek, artificial intelligence, optimization pathway

1. Introduction

Currently, the global financial system is at a critical juncture of digital transformation. Big data has emerged as the fourth major factor of production—following land, labor, and capital—profoundly reshaping the operational models and competitive landscape of the traditional financial industry. According to forecasts by the International Data Corporation (IDC), global data volume is projected to surge to 175 zettabytes (ZB) by 2025, with an annual growth rate of 40%. By 2026, the total data volume of Chinese enterprises is expected to exceed 100 ZB. As a highly data-intensive sector, the financial industry is likely to experience data growth rates above the average, implying that the scale of data in China's financial sector could approach or reach the 100 ZB level by 2026. Against this backdrop, the integration of artificial intelligence technologies with financial investment has become increasingly close, giving rise to the emerging field of intelligent financial investment. Its core characteristics are manifested in the high degree of synergy among data-driven approaches, algorithmic decision-making, and real-time execution.

In October 2025, during the "Alpha Arena" Global AI Trader Competition, the domestically developed DeepSeek large model stood out with outstanding performance. Under strict constraints on trading frequency (executing only 6 trades), the model achieved a cumulative return of up to 42% while demonstrating exceptionally strong risk control capabilities: its maximum single-trade loss was effectively limited to 348 USD, with an overall profit-loss ratio approaching 4:1 [1]. This milestone event not only demonstrated the formidable competitiveness of Chinese indigenous AI models in complex financial decision-making tasks but also clearly signaled that financial investment has officially transitioned from a human-dominated, experience-driven era to a data- and algorithm-centric intelligent-driven era.

The success of DeepSeek to some extent reflects China's long-term accumulation in financial technology, the continuous improvement of big data infrastructure, and ongoing innovations in algorithmic theory. At present, leading domestic financial institutions—such as public funds Huaxia Fund and Bosera Fund, as well as top brokerages Guotai Junan Securities and Zhongtai Securities—have successively completed the private deployment of intelligent models similar to DeepSeek and applied them to core business scenarios including intelligent investment research, automated risk control, and personalized asset allocation. According to the latest survey by the China Securities Industry Association, more than 70% of securities firms have incorporated artificial intelligence technologies into investment decision-making processes, and the assets under management by intelligent robo-advisors have surpassed 5 trillion RMB.

Nevertheless, the large-scale application of artificial intelligence in financial investment also introduces new challenges and risks. Algorithm homogenization may lead to "algorithmic stampede" events, exacerbating market volatility [2]. When a large number of market participants adopt intelligent algorithms with similar logic and converging data sources, strategy effectiveness can decay rapidly, potentially triggering vicious cycles in which strategies fail simultaneously. The high complexity of intelligent investment systems poses severe tests to traditional risk management frameworks, while the "black box" nature of algorithms makes risk tracing and effective regulation extraordinarily difficult. In addition, issues such as data privacy protection, algorithmic ethical responsibility, and technological monopolies are becoming increasingly prominent and urgently require systematic solutions.

Against this background, this study stands at the interdisciplinary frontier of financial engineering and data science, employing a combination of literature review, case analysis, and comparative research methods, with the following objectives: First, using representative cases such as DeepSeek as entry points, to deeply deconstruct the underlying technical logic and paradigm-shift characteristics that enable intelligent financial investment; second, to systematically review the current research progress and core bottlenecks in this field; third, taking into account the specific environment and development stage of the Chinese financial market, to construct a multi-layered, systematic optimization pathway for the development of financial investment.

2. Literature Review

Intelligent financial investment, as an emerging interdisciplinary research field, draws its theoretical foundation from the integration of multiple disciplines, including finance, computer science, statistics, and complex systems science. In recent years, scholars have conducted extensive discussions on its technical foundations, application progress, and potential issues.

2.1 Technical Foundations of Intelligent Financial Investment

Artificial intelligence technologies, represented by deep learning, serve as the core engine driving this round of transformation in financial investment. Domestically advanced models, exemplified by DeepSeek, typically adopt complex architectures such as Transformers and Mixture-of-Experts systems. Research indicates that these models, by incorporating dynamic routing or attention mechanisms, can adaptively activate the most relevant neural network modules or expert sub-networks based on varying input data features [1]. This design not only aligns theoretically with the nonlinear and non-stationary characteristics of financial data but also achieves a favorable balance between prediction accuracy and computational efficiency in practice. Yang Dong and Huang Tao point out that another major advantage of such models lies in their increasingly enhanced engineering and deployability, with relatively low costs for fine-tuning and private deployment, significantly lowering the barriers for financial institutions—especially small and medium-sized ones—to adopt cutting-edge AI technologies and accelerating the inclusive process of intelligent investment technologies [1]. However, most current studies remain at the level of macroscopic descriptions and performance demonstrations of model applications. There is still a lack of in-depth and detailed empirical testing and theoretical interpretation regarding the specific internal feature extraction mechanisms, the causal chains of decision formation, generalization capabilities and failure boundaries under different market regimes, such as bull, bear, and range-bound markets, as well as the intrinsic connections between model parameters and financial economic theories.

2.2 Progress in Investment Research Driven by Big Data

The rise of the intelligent investment paradigm is closely linked to the enormous enrichment of big data resources. Scholars' research perspectives are continuously expanding with the broadening of data dimensions. At the micro data structure level, distributed ledger technologies represented by blockchain naturally generate public, transparent, and tamper-proof transaction flow data. Wu Jiajing et al. argue that this provides researchers with an unprecedented fine-grained perspective for analyzing capital flows in cryptocurrency markets, identifying core holder address clusters, and characterizing complex transaction network topologies [3]. At the level of investment targets and strategies, emerging markets with high volatility and weak efficiency—such as digital finance and cryptocurrencies—have become natural "experimental fields" for testing the predictive capabilities of complex algorithms and exploring micro-market behaviors (such as herding effects and overreactions) [4]. Zhu Xiaoneng et al. (2025), in their monograph, systematically analyzed the investment logic and risk characteristics in these new financial markets, injecting fresh perspectives into traditional financial theories [4]. Nevertheless, a common deficiency in this field of research is the excessive reliance on historical backtesting. Many intelligent strategies that claim superior performance are constructed and optimized heavily based on local patterns from specific historical periods, lacking rigorous out-of-sample testing—particularly under extreme market stress scenarios not involved in training—and theoretical attribution analysis, which casts doubt on their robustness and long-term effectiveness.

2.3 Research Deficiencies and Innovation Opportunities

Synthesizing the existing literature, this study identifies three major shortcomings in current intelligent financial investment research that urgently need to be addressed, which also serve as the starting point and innovation space for this paper.

First, there is a systemic deficiency in the macroscopic theoretical framework. Existing studies are mostly distributed in a "point-like" or "linear" manner, focusing either on specific algorithm models or on certain data applications, failing to comprehensively outline—from a systems theory perspective—the profound impact on financial markets of the complete chain whereby artificial intelligence technologies proceed from underlying multi-source data fusion, to mid-level algorithmic decision-making, and then to macroscopic market structure evolution and cross-market risk contagion [5]. Intelligent finance constitutes a complex ecosystem that requires an integrated theoretical framework for understanding.

Second, at the micro-technical level, there exists a dilemma of algorithmic transparency and interpretability. The "black box" nature of deep learning models is the price paid for their powerful fitting capabilities. Huang Yiping sharply points out that this opacity not only hinders investors' understanding and trust in algorithmic decisions but also poses severe challenges to financial regulation: when regulators cannot effectively scrutinize core algorithmic logic, timely detection and suppression of potential market manipulation, algorithmic collusion, and other violations become extraordinarily difficult [2]. Breaking the "black box" to achieve trustworthy and controllable intelligent decision-making is key to the healthy development of the industry.

Finally, at the level of application in specific national contexts, there is a lack of localization and comprehensiveness in research perspectives. Many studies draw on overseas models and cases but give insufficient consideration to the unique institutional environment, investor structure, and data characteristics of China's multi-layered and multi-segment capital markets. Li Zhan's in-depth research on the investor structure and risk control in China's capital markets provides an important foundation for understanding the complexity of the domestic market [6]. Future research needs to pay greater attention to analyzing the cross-contagion mechanisms of risks potentially triggered by intelligent algorithmic trading across different markets such as A-shares, bond markets, derivatives, and digital assets, in order to construct risk control systems suited to national conditions.

3. Paradigm Shift in Intelligent Financial Investment

The outstanding performance of DeepSeek in real-world practice is not the result of a single technological breakthrough, but rather represents a comprehensive paradigm shift encompassing data, models, and risk control.

3.1 Data Dimension: From Structured to Multi-Source Heterogeneous Fusion

Traditional quantitative investment heavily relies on highly structured data such as historical prices, trading volumes, and financial indicators. While such data is well-organized and clean, its informational dimensions are relatively limited and exhibit clear lags. In contrast, the intelligent investment paradigm has achieved a fundamental transformation toward the fusion of multi-source heterogeneous big data. Investor sentiment extracted from social media texts, port logistics and factory activities captured by satellite imagery, IoT data from upstream and downstream supply chains, real-time transaction flows on blockchain, and more are all incorporated into the decision-making framework. Empirical research by Chen Hua and Chu Wenrong demonstrates that automated data pipelines constructed using natural language processing and computer vision technologies can convert unstructured information into quantitative signals with an accuracy rate exceeding 92% [7]. This has revolutionized the “information boundary” of investment decision-making.

However, multi-source data fusion also introduces new challenges, including varying data quality, inconsistent standards, and differences in timeliness, necessitating the development of more intelligent data cleaning and fusion algorithms. Conflicts may arise between different data sources—for instance, sentiment indicators from social media may send signals completely opposite to those from fundamental data. How to handle these conflicting signals and how to assign appropriate weights to different data sources have become critical issues in the design of intelligent investment systems.

3.2 Model Algorithms: From Linear Analysis to Nonlinear Deep Learning

Traditional quantitative models are predominantly based on linear regression, the Capital Asset Pricing Model (CAPM), mean-variance frameworks, and other linear or parametrically nonlinear assumptions. These models often fall short in capturing the widespread complex nonlinear interactions, dynamic time-varying relationships, and long-range dependencies present in financial markets. Models represented by DeepSeek, leveraging deep neural networks, attention mechanisms, and reinforcement learning technologies, have achieved a leap from a “human hypothesis–data validation” paradigm to a “data-driven–automatic discovery” mode.

Wang Lei and Liu Jing point out that such models can automatically mine high-dimensional, nonlinear features and patterns from massive datasets and continuously optimize themselves based on dynamic environmental feedback, demonstrating stronger adaptability and predictive power in high-noise, high-volatility markets [8]. Nevertheless, the complexity of deep learning models also brings risks of overfitting, poor interpretability, and high training costs. Particularly when market structures undergo fundamental changes, models trained on historical data may fail to adapt to new market environments, resulting in the failure of investment strategies. The massive market shocks triggered by the 2020 COVID-19 pandemic exposed the vulnerability of many intelligent models, as these models had never encountered such extreme market conditions in their training data, resulting in a sharp decline in their predictive and decision-making capabilities.

3.3 Risk Control System: From Static Thresholds to Dynamic Scenario Awareness

Traditional risk control systems typically rely on static risk metrics calculated through historical data backtesting and periodic evaluation frequencies. This approach responds slowly and fails to distinguish between different market scenarios, rendering it prone to fail in extreme conditions. The core of the new generation of intelligent risk control systems is “dynamic scenario awareness.” Relying on streaming computation engines and complex event processing technologies, these systems enable millisecond-level real-time monitoring of investment portfolios.

More importantly, by integrating reinforcement learning and simulation technologies, the system can construct a virtual “parallel market” in which thousands of path simulations are performed on the current portfolio under various market scenarios. On this basis, the system can not only dynamically compute real-time risk values but also proactively and predictively adjust risk mitigation measures. This marks the evolution of risk control from a passive, rule-based "threshold alarm" to an active, prediction-driven "intelligent autopilot."

4. Optimization Pathways for the Development of Financial Investment

To address the opportunities and challenges brought by this paradigm shift and to promote the healthy and steady intelligent transformation of the financial investment industry, this study constructs a systematic optimization framework comprising five dimensions.

4.1 Building a Secure and Efficient Data Factor Circulation Ecosystem

Data serves as the "oil" of intelligent finance, yet its circulation and application face numerous obstacles such as unclear ownership rights, privacy leakage risks, and pricing difficulties. The first step in the optimization pathway is to construct a "trustworthy data infrastructure." On the technical path, active exploration should be undertaken to integrate blockchain technology with privacy-preserving computing techniques such as federated learning and secure multi-party computation. Blockchain can provide an immutable ledger for data ownership verification, access logging, and transaction tracing, while technologies like federated learning enable data to be "usable yet invisible," fundamentally resolving the contradiction between data privacy and collaborative computation.

In terms of application scenarios, efforts should be made to promote the compliant integration of cross-domain data—such as meteorology, environment, and logistics—with financial data. This will offer unprecedented multi-dimensional perspectives for assessing macroeconomic trends and industry prosperity. For example, combining meteorological data with agricultural insurance product design can lead to the development of more precise weather derivatives; integrating logistics data with supply chain finance can enable real-time evaluation of supply chain health and early warning of potential risks.

4.2 Enhancing Transparency and Robustness of Intelligent Algorithms

The reliability and trustworthiness of algorithms form the cornerstone of the long-term development of intelligent finance. This requires coordinated advancement at both the model design and external regulatory levels. At the model design level, priority should be given to developing algorithms capable of fusing data across different frequencies, thereby integrating macroeconomic cycle signals, meso-level industry signals, and micro-level trading signals. At the same time, causal inference frameworks should be actively introduced to advance models from correlation mining to causal relationship exploration, thereby enhancing the logical interpretability of strategies.

At the regulatory coordination level, the mandatory adoption of explainable artificial intelligence techniques in core models must be enforced, requiring visual and logical attribution analysis for key decisions. Regulatory authorities should also establish mechanisms for assessing the diversity of algorithmic strategies to prevent systemic resonance risks arising from strategy homogenization in the market. Furthermore, routine adversarial sample testing and extreme stress scenario simulation exercises should be conducted to ensure algorithmic stability under extreme market conditions.

4.3 Elastic Transformation of Market Structure

The widespread adoption of intelligent algorithms is reshaping market microstructures, and market infrastructure itself must undergo adaptive evolution to enhance resilience. At the trading mechanism level, exchanges may consider rule innovations. For instance, introducing order types or fee structures that reward the provision of differentiated and genuine liquidity, as well as offering fee reductions or other incentives to algorithmic strategies that continue to provide liquidity during periods of market stress, can institutionally prevent the risk of sudden liquidity evaporation caused by synchronized algorithmic withdrawals due to homogenization.

At the technical support level, regulatory authorities should take the lead in building market monitoring platforms based on multi-agent simulation. Such platforms can simulate interactions among different strategies in virtual markets, thereby enabling real-time early warning and stress testing for potential liquidity dry-up points and cross-market risk contagion. These simulation platforms can become an important component of the regtech ecosystem, helping regulators transition from a "post-event handling" to a "pre-event foresight" regulatory paradigm.

4.4 Reconstructing a Regulatory System Adaptive to Technological Development

In the face of technology-driven financial innovation, the regulatory system requires greater intelligence, stronger flexibility, and more advanced tools. In terms of institutional design, the regulatory sandbox regime should be further improved and promoted, providing a risk-isolated real-world testing environment for genuinely innovative fintech applications and establishing a dynamic balancing mechanism between encouraging innovation and preventing risks. China's fintech regulatory sandbox pilots have already achieved initial success; the next step should be to expand the pilot scope, refine evaluation and promotion mechanisms.

In terms of regulatory capacity building, regtech must be vigorously developed. This means that regulatory authorities themselves need to deeply employ big data, artificial intelligence, and other technologies for self-empowerment—using natural language processing to automatically review massive volumes of disclosure texts, and knowledge graph techniques to map complex networks of related-party transactions—thereby enhancing penetrating and full-coverage supervisory capabilities over modern financial markets.

4.5 Upgrading a Full-Cycle Intelligent Risk Management System

The risk management system must evolve in sync with the development of intelligent investment—and even be deployed ahead of time. At the cutting edge of technological application, the construction of "digital twin" risk control systems for investment portfolios can be explored. By synchronizing real portfolio data in real time, these systems perform millisecond-level scenario deductions and stress tests in a virtual space. This approach provides more comprehensive and in-depth risk assessments for analyzing extreme yet possible "black swan" events.

In terms of risk governance architecture, a three-tier collaborative defense system—micro, meso, and macro—should be established. At the micro level, the principal responsibility of financial institutions should be reinforced by requiring them to set up independent model validation teams. At the meso level, exchanges, industry associations, and similar organizations should establish overall market algorithmic strategy databases and crowding indices. At the macro level, regulatory authorities need to research and formulate counter-cyclical adjustment tools targeted at algorithmic trading, mitigating systemic vulnerabilities from a macroprudential perspective. Notably, intelligent risk management should not completely replace human judgment. At the most critical risk decision points, human override mechanisms must be retained to ensure human intervention and final decision-making in the event of system anomalies or extreme situations.

5. Conclusion

This study systematically dissects the profound paradigm shift occurring in the field of financial investment under the impact of the big data and artificial intelligence technology wave, revealing the core characteristics from data fusion and algorithm evolution to risk control upgrading. On this basis, it innovatively proposes a systematic optimization framework encompassing five dimensions: data ecosystem, algorithm governance, market structure, regulatory system, and risk management. The framework emphasizes that the intelligent transformation of financial investment is a complex systems engineering project involving technology, institutions, markets, and regulation. Only through multi-party collaboration and comprehensive measures—while encouraging technological innovation and at the same time strengthening the dam of risk prevention and control—can we guide China's financial technology toward steady and far-reaching development, truly enhancing the efficiency and quality of financial services in support of the real economy.

The development of intelligent financial investment in China possesses unique advantages and challenges. Its advantages lie in the enormous market size, abundant data resources, leading digital infrastructure, and strong policy support; the challenges include relatively low market maturity, a distinctive investor structure, and restrictions on cross-border capital flows. Therefore, the development path of intelligent financial investment in China should not simply copy foreign models but should explore an independent innovation route that aligns with national conditions and characteristics.

Looking ahead, relevant research and practice still need to deepen exploration in the following directions: First, research on algorithmic robustness under extreme market conditions, particularly in the Chinese market environment characterized by frequent policy interventions and active retail trading, where robustness testing must pay greater attention to indigenous features. Second, assessment of the ethical and social impacts of intelligent finance, necessitating the establishment of corresponding ethical guidelines to prevent algorithmic discrimination and protect the rights and interests of financial consumers. Third, deep integration of financial theory and artificial intelligence theory, leveraging AI technologies to discover new financial regularities and develop new financial theories adapted to the intelligent era.

It is foreseeable that, with continuous technological iteration and ongoing institutional improvement, intelligent financial investment will evolve toward greater efficiency, transparency, and resilience. This process is not only a self-renewal of the financial industry but will also contribute Chinese wisdom and Chinese solutions to the stability and innovation of the global financial system.

References

[1] Yang, D. and T., H. Deepseek-like artificial intelligence empowering fintech 3.0: Systemic transformation and regulatory adaptation. Journal of Financial Research. 2025(3), pp. 45-48.

[2] Huang, Y. P. Financial regulation in the era of artificial intelligence. Economic Research Journal. 2025(5), pp. 78-92.

[3] Wu, J. J., Li, G. Q. and Sun, Y. A review of research on blockchain transaction networks. Computer Science. 2021, 48(9), pp. 12-24.

[4] Zhu, X. N., Shan, C. Y. and Y., X. H. Chinese financial market studies: New financial investment theory and practice. Shanghai: Shanghai University of Finance and Economics Press, 2025.

[5] Zhu, X. N., Kang, W. J. and X., Z. X. Principles of Investment and China's Financial Markets. 2nd ed. Shanghai: Shanghai University of Finance and Economics Press, 2024.

[6] Li, Z., Tang, J. R. and Yao, Y. Z. The real logic of capital markets: Research on investor structure, wealth effects and risk control in China's capital markets. Beijing: Economic Management Press, 2022.

[7] Chen, H. and Chu, W. R. An empirical study on deepseek in intelligent decision-making of commercial banks. Financial Forum. 2025(4), pp. 33-45.

[8] Wang, L. and Liu, J. Progress and challenges of deep learning in quantitative investment. Statistics and Decision. 2024, 40(8), pp. 89-94.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.