Analysis on the Application Value and Practical Path of XAI in Chinese Intelligent Finance

Yiqi Dong
School of Business (BBA),Macau University of Science and Technology, Macao, China

Abstract: With the wide application of artificial intelligence in China's financial industry, the AI black-box problem poses multiple risks to regulation, risk control and consumer protection. Research shows that the application of XAI can effectively improve the transparency and interpretability of financial AI and solve the black-box problem. Taking this as the starting point, based on sorting out relevant theories and regulatory practices at home and abroad, this paper puts forward the XAI promotion path from platform construction, institutional guarantee and talent training, and analyzes its application prospects via SWOT method. Results show that XAI has high application value but faces limitations in technology, governance, and talents. China's AI governance framework needs improvement in standards and supervision. Developing user-friendly modeling platforms, perfecting regulation, and cultivating interdisciplinary talents will support the sustainable development of XAI in finance.It provides a reference for financial institutions and regulators to jointly promote the implementation of XAI and ensure the safe and compliant development of intelligent finance.

Keyword: Black Box; XAI; Explainable AI

1. Introduction

Artificial intelligence (AI) has been widely applied in China's financial industry. As a typical data-intensive sector, finance, with its vast amount of high-quality data assets and diverse business scenarios, has become one of the areas where AI technology is most maturely applied and deeply integrated. With the transition of "AI + Finance" from concept to large-scale implementation, the penetration rate of large models in the financial sector exceeded 50% in 2024, with 63 publicly disclosed large model winning projects throughout the year and a total contract value exceeding 3.6 billion yuan [1].The application of AI is no longer limited to auxiliary roles in traditional financial service scenarios such as automated trading through rule-driven and expert systems. Rather, through methods such as machine learning [2] and the integration of big data and blockchain, AI has comprehensively covered the front, middle, and back offices of the financial industry, including core scenarios such as robo-advisory, intelligent risk control, and market analysis [1,2,3,4]. For example, in personal consumer credit approval, AI systems conduct credit assessments for online shopping installment loans, travel loans, and household appliance loans. AI is also widely used in transaction fraud detection, monitoring real-time fund transfers, card swipes, and online payments to spot abnormal transactions such as stolen card transactions and fake payments.As technology becomes increasingly complex, while AI empowers the financial industry, it also brings corresponding risks and challenges. Among them, the lack of transparency and explainability in the AI decision-making process, known as the "black box problem," has become increasingly prominent in China's financial scenarios. To address this, explainable artificial intelligence (XAI) has developed rapidly. Its central goal is to make the behavior of these systems understandable to humans by elucidating the underlying mechanisms of their decision-making processes [5].

This paper takes the AI black box problem as its starting point, focuses on XAI as a solution, and seeks to answer the following questions: (1) What is the concept, causes, and solution path of the AI black box problem? (2) How can XAI be effectively applied in Chinese financial practice? The academic community abroad has achieved remarkable results. DARPA officially proposed the XAI research agenda in 2016 [6]. Nowadays, the XAI technology classification system has become increasingly mature [7], and methods have diversified [8], including model-specific technologies such as decision trees [9], as well as post-hoc explanation methods like SHAP, LIME, Anchors [8], and XGNN [6]. Among these, representative methods such as SHAP [10], LIME, and counterfactual explanations have been widely applied in financial scenarios including credit approval, risk management, and fraud detection [11]. At the regulatory level, the EU's "Artificial Intelligence Act" focuses on rights protection and imposes mandatory transparency requirements for high-risk AI systems [12]; the United States adopts a market-driven regulatory approach, encouraging innovation while strengthening security testing and transparency constraints for high-risk AI [13]. China has implemented regulations such as the "Regulations on the Management of Deep Synthesis of Internet Information Services" [14] to achieve scene regulation and progressive system construction, thereby establishing a preliminary regulatory framework for algorithms and artificial intelligence [15]. However, China's current algorithm filing system faces issues such as insufficient effectiveness, incomplete rules, and lack of supporting measures [16], and lacks targeted and operational guidance. Overall, China still lags behind the international level in both the practice and theoretical research of XAI.

To address the aforementioned issues, this paper Starts from the AI black-box problem in the financial sector, analyzing the causes of black-box models and the technical system of Explainable Artificial Intelligence (XAI). Integrated with domestic and international regulation and practice, it explores the application path of XAI in China’s financial industry and evaluates its application prospects through SWOT analysis. This research enriches the study of XAI in financial scenarios, provides references for financial institutions to address algorithmic opaqueness, comply with regulatory requirements and protect consumer rights, and promotes the safer, more trustworthy and compliant development of financial AI.

The following literature review chapter defines the AI black box and XAI, analyzes the causes and challenges of the black box problem in finance, and sorts out the technical methods, application status and regulatory policies of XAI worldwide. The methodology chapter constructs the implementation path of XAI in China’s financial industry from three aspects: platform development, institutional guarantee and talent cultivation. The discussion chapter carries out a SWOT analysis to comprehensively evaluate the XAI application in China’s financial sector. The conclusion chapter summarizes the main conclusions of this paper and prospects the future development of explainable AI in the financial field.

2. Literature Review

2.1 Black Box

As shown in Figure 1, AI models can be categorized into white-box, grey-box, and black-box models based on transparency. White-box models are fully transparent; grey-box models balance complexity and clarity [6]; black-box models are highly opaque with uninterpretable internal mechanisms. This lack of interpretability gives rise to the “black box problem” in AI, where complex algorithms (especially deep learning) process inputs through hidden mechanisms to produce outputs [17]. Many state-of-the-art models, including large language models like ChatGPT and Llama, exhibit black-box characteristics [18]. While they achieve high accuracy, their opacity raises risks in critical applications where decision rationale must be clear [6].

The black box problem stems from two main causes. First, developers may deliberately hide internal workings to protect intellectual property, even in rule-based systems. Second, and more critically, deep learning models are inherently black boxes due to their extreme complexity-multi-layered neural networks with hundreds or thousands of layers and numerous neurons are so intricate that even creators cannot fully trace the internal processes that produce outputs [18].

This black-box nature poses severe challenges to the financial industry. At the operational level, it obscures data processing and transaction decisions, hindering internal risk controls and regulatory penetrating supervision. This makes risk tracing, management, and accountability difficult, leading to compliance and regulatory challenges [19,20]. At the level of derivative risks, the black box further spawns algorithm convergence, bias, and hegemony [20]. When multiple institutions adopt similarly flawed models, imperceptible common problems may cause systemic risks to accumulate across institutions [19]. At the consumer level, scholars note that the black box problem may “seriously infringe upon the interests and confidence of financial consumers” [21].

Given the seriousness of the financial sector, regulatory authorities have explicitly required that AI-driven financial decisions be interpretable, accountable, and fair. Thus, revealing the causes of the black box problem and exploring countermeasures are essential for preventing digital financial risks. In this context, Explainable Artificial Intelligence (XAI) has emerged as a key concept to address these challenges [20].

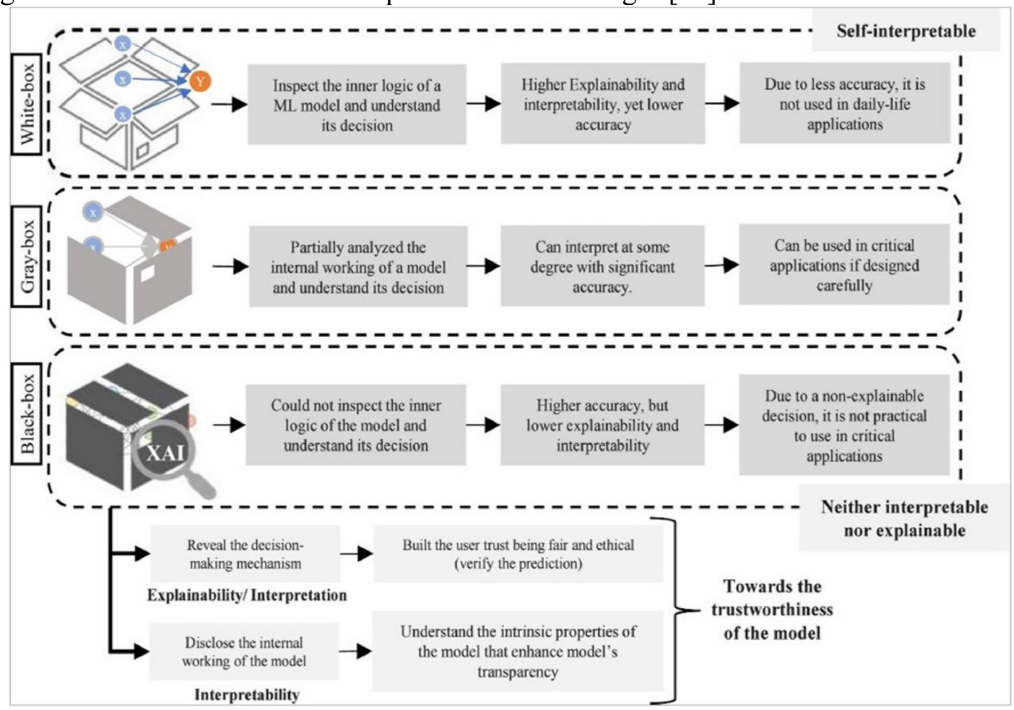

Figure 1. Illustration of White, Gray, and Black Box Models, Highlighting Their Transparency Levels in AI Systems[6]

2.2 XAI

Explainable Artificial Intelligence (XAI) refers to the ability of AI systems to provide clear and understandable explanations for their decisions, and effectively manage machine learning outcomes by characterizing model accuracy, fairness, transparency, and potential biases, thereby helping to build organizational trust in AI deployment [5,22]. As shown in Figure 2, XAI has distinct application scenarios and practical value in model decision-making. Its core goal is to make the decision-making process of AI models understandable to humans along the dual dimensions of explainability and interpretability, where interpretability is specifically reflected in the two principles of transparency and completeness.

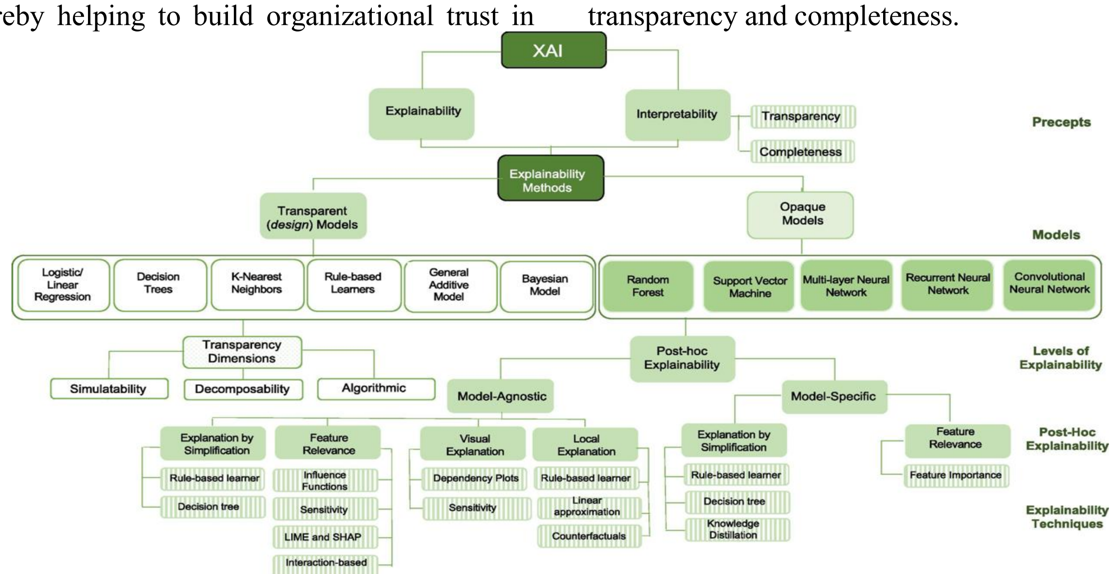

Figure 2. Conceptual Taxonomy of XAI[7]

Based on this objective, XAI divides existing models into two categories: the first is transparent design models, such as decision trees, and logistic regression, . These models have a clear structure and can directly demonstrate their decision logic through the three dimensions of simulatability, decomposability, and algorithmic interpretability. Internationally, such models have long become benchmark methods in the financial field. For example, European and American banks widely use logistic regression and decision tree rule-based models in credit approval, credit scoring, and default prediction. Their interpretation process is intuitive and easy to audit, meeting the compliance requirements of global financial regulation for a long time [23-24]. The second category consists of complex opaque models such as random forests and support vector machines. These models perform prominently in financial tasks including credit risk assessment, stock price prediction, anti-fraud, and anti-money laundering, but their internal mechanisms are difficult to interpret directly. Therefore, post-hoc explainability techniques are needed to decompose their decision logic. Internationally, Soori et al. applied XAI to financial tasks such as stock price prediction, and credit assessment. Ballegeer et al. verified the stability of SHAP and LIME in credit scoring [17]. In China, Duxiaoman has developed and deployed a financial risk control XAI system that enables interpretable credit decisions at both global and individual transaction levels, while generating natural language explanations [25], and Everbright Securities combined large models with knowledge graphs to detect suspicious transactions [26]. All of these applications adopt model-agnostic methods such as feature correlation analysis and visual interpretation, as well as model-specific methods such as knowledge distillation, which together form a complete technical system of XAI in the financial field that meets regulatory requirements.

With the spread of XAI, regulatory policies have emerged. The EU AI Act (2024) classifies AI by risk: unacceptable risks banned; high-risk (e.g., credit approval) strictly supervised; limited-risk requires transparency; minimal-risk unrestricted [27]. The U.S. Biden executive order (2023) was revoked in 2025, replaced by a new order reducing restrictions and granting more autonomy. The SEC and CFPB have appointed chief AI officers [28]. China's Interim Measures for Generative AI (2023) adopt inclusive, graded supervision, requiring algorithm transparency, filing, and security assessment. In finance, the PBOC leads governance; the Financial Technology Development Plan (2022–2025) promotes intelligent technologies; the Ethical Guidelines for AI Applications in the Financial Field (2022) clarify ethical principles [21].

3. Methodology

Based on a review of explainable AI (XAI) applications in China's financial industry, the implementation of XAI can be systematically discussed from three dimensions: platform construction, institutional guarantee, and talent development.

3.1 Platform Construction

To lower technical barriers, a user-friendly unified modeling platform is essential. A visual decision-path tracing mechanism (e.g., flowcharts showing key nodes and activation intensity) helps business staff understand complex models like deep learning. In financial risk control, this makes the "reasoning chain" behind transaction blockages visible. Additionally, simple rapid-modeling components-such as drag-and-drop, pre-built modules, and visual configuration-allow non-technical analysts to perform data access, model training, and process orchestration without coding. This boosts efficiency, unites tech and business teams, and enables widespread modeling capabilities.

3.2 Institutional Guarantee

An industry-wide certification system and stronger regulatory technology are needed. For example, CAICT has launched the "Trustworthy AI"interpretability grading certification, providing quantifiable evaluation criteria [29]. Following this, efforts should fill gaps in interpretability standards, turning XAI from a "differentiated advantage" into an "access threshold." Regulation is equally critical. As former CBRC chairman Shang Fulin noted, regulators must enhance penetrating analysis of AI algorithm risks, improve relevant rules, and ensure algorithm interpretability, transparency, fairness, and security[30]. This means regulators themselves need the technical capability to effectively examine financial institutions' AI models.

3.3 Talent Development
Strengthening interdisciplinary talent training is vital. The adoption of XAI in finance requires professionals skilled in financial business, risk management, AI technology, and algorithm interpretation. Financial institutions should provide targeted practical training for technicians, risk controllers, business managers, and auditors. Moreover, collaboration among financial institutions, universities, and research institutes should be enhanced to develop training models tailored to real-world financial applications, ensuring sufficient professional talent for the sustainable implementation of explainable AI.

4. Discussion
This section provides a SWOT analysis of the large-scale application of explainable artificial intelligence (XAI) in China's financial industry.

4.1 Strengths
Domestic AI technologies and visualization tools have developed rapidly, and XAI has become a major research hotspot and key development direction in artificial intelligence. Driven by industry-wide progress, model explanation, decision traceability, and visualization tools are continuously optimized to clearly demonstrate the judgment basis and reasoning process of AI models. Meanwhile, financial institutions have undergone in-depth digital transformation and possess relatively mature data platforms, AI modeling capabilities, and engineering implementation conditions. Together, these factors provide solid and stable support for the application and large-scale promotion of XAI in the financial industry.

4.2 Weakness
Current technical capabilities are limited, and the interpretation effect and stability of complex AI models remain insufficient. Meanwhile, there is a severe shortage of interdisciplinary talents who master both financial business and AI model explanation, which restricts the large-scale application of XAI in the financial industry.And the development of XAI-related technologies, model modification and system deployment require relatively high investment, and the cost of later maintenance and continuous optimization is also high, which further increases the difficulty of its large-scale implementation in financial institutions.

4.3 Opportunity
Governments and financial regulators have vigorously promoted algorithmic transparency and explainable AI. For instance, in early 2026, the National Administration of Financial Regulation issued the Guiding Opinions on Regulating the Application of Artificial Intelligence Models in Financial Institutions, which states that "automated decision-making analysis, model algorithm development, data annotation and other activities shall ensure the transparency of data processing and the fairness and rationality of results", requiring models to be explainable, verifiable, traceable and auditable. This has provided clear policy guidance and compliance impetus for the development and implementation of XAI. Meanwhile, mature global experience and research achievements in XAI can be drawn upon to reduce trial costs in domestic exploration and application. For example, the EU's AI Act has established clear mandatory requirements for the explainability of high-risk financial AI systems, and its mature regulatory and technical practices offer valuable references for local applications[27].

4.4 Threat
AI models applied in the financial sector are becoming increasingly complex, which demands more sophisticated and specialized explainability methods. However, existing technologies can hardly fully restore their decision-making logic, leading to a significant rise in the difficulty and cost of interpretation. Meanwhile, there is an inherent contradiction between model transparency and data privacy protection. XAI requires transparency in the model decision-making process, but financial data involves a great deal of customer privacy and trade secrets. Improving explainability often means exposing more model logic and data details, thereby increasing the risks of algorithm reverse engineering, fraudulent attacks and privacy breaches.

5. Conclusion
The AI black box problem refers to the lack of transparency and interpretability in complex AI decision-making. Its main causes are the structural complexity of deep learning models and their opaque internal logic. This problem

can be addressed through the development of Explainable AI (XAI), which enhances model explainability and improves algorithmic transparency. Currently, XAI has been gradually implemented in China's financial sector across scenarios such as credit risk control, fraud detection, and robo-advisory. By improving the transparency of model decision-making, it helps financial institutions meet regulatory requirements and enhance user trust.While XAI holds significant application value in the financial field, it still faces limitations in terms of technical stability, computational efficiency, and institutional governance. Regarding the application of XAI, although China has initially established an AI governance framework, it still needs further improvement in terms of interpretation standards and regulatory coordination. In the future, XAI can be effectively applied in Chinese financial practice by building user-friendly modeling platforms that lower technical barriers and empower non-technical staff to model efficiently. At the same time, establishing standardized regulatory and certification systems, along with strengthening interdisciplinary talent development, will support the sustainable adoption of XAI. Through these efforts, interpretable models can be deployed for effective application in China’s financial practice.

Reference

[1] Huang, L., & Zhang, H. (2026). Evolution of Large Chinese Financial Models in the Context of Artificial Intelligence: *Journal of Global Information Management*, *34*(1). https://doi.org/10.4018/JGIM.402742

[2] Aldasoro, I., Gambacorta, L., Korinek, A., Shreeti, V., & Stein, M. (2025). Intelligent financial system: How AI is transforming finance. *Journal of Financial Stability*, 101472. https://doi.org/10.1016/j.jfs.2025.101472

[3] Yan, L. S., Fan, K. H., Wang, Z., & Zhu, B. H. (2024). *The development history and prospects of AI in the financial industry* [J].Hanspub.org. https://www.hanspub.org/journal/paperinformation?paperID=101683#return20

[4] Zhou, H. T. (2025). A*pplication and governance of artificial intelligence in the financial industry*. Jin Rong Zong Heng, (9),68–74.https://kns.cnki.net/kcms2/article/abstract?

[5] *TechDispatch 2/2023 - Explainable Artificial Intelligence | European Data Protection Supervisor*. (2024, March 15). www.edps.europa.eu. https://www.edps.europa.eu/data-protection/our-work/publications/techdispatch/2023-11-16-techdispatch-22023-explainable-artificial-intelligence_en

[6] Mohamed, A., Abdelqader, K., & Shaalan, K. (2025). Explainable Artificial Intelligence: A Systematic Review of Progress and Challenges. *Intelligent Systems with Applications*, 200595. https://doi.org/10.1016/j.iswa.2025.200595

[7] Peter E.D. Love, Fang, W., Matthews, J., Porter, S., Luo, H., & Ding, L. (2023). Explainable Artificial Intelligence (XAI): Precepts, models, and Opportunities for Research in Construction. *Advanced Engineering Informatics*, *57*, 102024–102024. https://doi.org/10.1016/j.aei.2023.102024

[8] Ali, S., & Abuhmed, T. (2023). Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. *Information Fusion*, *99*(101805), 101805. sciencedirect. https://doi.org/10.1016/j.inffus.2023.101805

[9] Christoph Molnar. (2019, August 27). *Interpretable Machine Learning*. Github.io. https://christophm.github.io/interpretable-ml-book/

[10] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30.https://doi.org/10.48550/arXiv.1705.07874

[11] Bussmann, N., Giudici, P., Marinelli, D. et al. Explainable Machine Learning in Credit Risk Management. Comput Econ 57, 203–216 (2021). https://doi.org/10.1007/s10614-020-10042-0

[12] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on Artificial Intelligence (AI Act). Official Journal of the European Union.https://eur-lex.europa.eu/eli/reg/2024/1689/oj

[13]Executive Order 14110. (2023, October 30). Safe, secure, and trustworthy development and use of artificial intelligence. Federal Register, 88(210), 74743–74776. https://www.federalregister.gov/d/2023-24283

[14] Cyberspace Administration of China, Ministry of Industry and Information Technology, & Ministry of Public Security. (2022). *Provisions on the Administration of Deep Synthesis of Internet Information Services (Order No. 12 of the Cyberspace Administration of China and Other Departments*). Cyberspace Administration of China.https://www.cac.gov.cn/2022-12/11/c_1672221949354811.htm

[15] China Institute for Legal Society, Shanghai Jiao Tong University. (2025). Sjtu.edu.cn. http://www.socio-legal.sjtu.edu.cn/wxzy/info.aspx?itemid=4843

[16]Li, J. J. (2025). *A review of the current situation and path exploration of generative artificial intelligence algorithm filing*. Journal of University of Electronic Science and Technology of China (Social Sciences Edition), 27(5), 109–117.https://doi.org/10.14071/j.1008-8105(2025)-5016

[17]Blouin, L. (2023, March 6). *AI's mysterious "black box" problem, explained*. University of Michigan-Dearborn. https://umdearborn.edu/news/ais-mysterious-black-box-problem-explained

[18]Kosinski, M. (2024, October 29). *Black box AI*. Ibm.com. https://www.ibm.com/cn-zh/think/topics/black-box-ai

[19]Han, L., & Liu, Y. (2025). *Challenges posed by financial digitalization to financial regulatory objectives and corresponding countermeasures*. Sufe.edu.cn. https://qks.sufe.edu.cn/mv_html/j00003/202504/wfsHgnb7-dhhe-zdu8-ImCU-7FpAPHjqqAwg_WEB.htm

[20]Yildiz, K., & Cil, A. E. (2025). A systematic literature review on applications of explainable artificial intelligence in the financial sector. *Internet of Things*, 101696. https://doi.org/10.1016/j.iot.2025.101696

[21]Mo, W. G., & Xu, R. H. (2025). *Risks and responses to the application of artificial intelligence in the financial sector*. Xinhua Finance - China Financial Information Network.https://www.cnfin.com/hg-lb/detail/20251117/4336061_1.html

[22]IBM. (2023, March 29). *Explainable AI*. IBM. https://www.ibm.com/think/topics/explainable-ai

[23]Lessmann, S., Baesens, B., Seow, H.-V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*, *247*(1), 124–136.https://www.sciencedirect.com/science/article/pii/S0377221715004208

[24]Dastile, X., Celik, T., & Potsane, M. (2020). Statistical and machine learning models in credit scoring: A systematic literature survey. *Applied Soft Computing*, *91*, 106263. https://doi.org/10.1016/j.asoc.2020.106263

[25]Zhongxin Jingwei (Economic View). (2026). *What Have Leading Fintech Companies Been Doing in the Past Year?*. (2026). Jwview.com. https://www.jwview.com/jingwei/html/m/01-17/654758.shtml

[26]Shanghai Stock Exchange. (2024). *Notice on launching the 2024 industry joint research projects of the Securities Information Technology Research and Development Center (Shanghai)*. SSE.com.cn.https://www.sse.com.cn/services/tradingtech/notice/c/c_20241122_10764876.shtml

[27]European Union. (2025). *The EU Artificial Intelligence Act*. The Artificial Intelligence Act. https://artificialintelligenceact.eu/

[28]tshaw. (2025, August 5). *SEC Launches AI Task Force to Drive Innovation and Efficiency*. Thomson Reuters Tax & Accounting News. https://tax.thomsonreuters.com/news/sec-launches-ai-task-force-to-drive-innovation-and-efficiency/

[29]Artificial Intelligence Research Institute of China Academy of Information and Communications Technology, Computational Social Science and National Governance Laboratory of Tsinghua University, & Data Committee of China Artificial Intelligence Industry Development Alliance. (2025). *Guidelines for the Construction of High-Quality Artificial Intelligence Datasets*. https://www.caict.ac.cn/kxyj/qwfb/ztbg/202508/P020250829553828220259.pdf

[30]Shang, F. L. *Risks in the financial industry, such as black-box AI algorithms, data contamination, and model defects, cannot be ignored*.(2025). Caixin.com. https://finance.caixin.com/2025-04-11/102308238.html

