Proceedings of ICFTBA 2025 Symposium: Data-Driven Decision Making in Business and Economics
DOI: 10.54254/2754-1169/2025.BL29505

AI + Big Data: Transforming Traditional Credit Reporting

Yuchi Chen[1], Yifei Pei[2*]

[1]College of Accounting, Beijing Normal University-Hong Kong Baptist University, Zhuhai, China
[2]College of International Education Cooperation, Tianjin University of Commerce, Tianjin, China
**Corresponding Author. Email: peiyifei@stumail.tjcu.edu.cn*

Abstract. Against the background of a traditional credit system that can no longer meet today's needs, this paper systematically analyzes its structural shortcomings, including one-dimensional data sources, outdated information updates, and severe data fragmentation, which explains the reason that it hampers inclusive-finance development. On this basis, it then explores how artificial intelligence and big-data technologies can be applied to credit reporting through three key relationship-network construction, including the use of relationship networks, full-dimension credit profiling, and dynamic risk alerts to improve both assessment accuracy and real-time timeliness. The study further highlights multiple risks in AI-driven credit systems, such as legal compliance challenges, privacy leakage, and model explainability risks inherent in AI-driven credit scoring. Accordingly, it proposes optimizing and governing a multi-level governance framework through the three layers: relationship networks, full-dimension profiling, and dynamic alerts, to push the industry toward greater precision, inclusiveness, and a secure development path for the credit industry.

Keywords: Traditional credit reporting, AI credit reporting, Big data, Credit assessment, Risk prediction

1. Introduction

The traditional credit-reporting regime, anchored in the central-bank credit file, is trapped by one-dimensional data and slow updates, leaving "thin-file" consumers and micro-entrepreneurs outside the fence and developing information islands. Nowadays, the rise of artificial intelligence and big data makes it possible to fold unstructured signals—social-media footprints, e-commerce transactions, public financial statements—into credit assessments, enabling dynamic risk discovery and a new leap for the industry.

The existing research already confirms the value of AI injection in the macro level: Cheng Wengai & Li Rui show that adding e-commerce cashflow data sharply lifts default prediction accuracy in the grain-and-oil sector [1]; Doğuş Emin et al. use the GMM system on OECD and BRICS panels to prove that AI scores outperform traditional scorecards [2]; Wenqing Bao et al. conducted a comparative study on credit risk management in real estate supply chain finance based on GA-SVM, providing a comprehensive evaluation of AI models versus traditional models and further quantifying the marginal contributions of AI models [3]. Yet these studies mostly focus on

DOI: 10.54254/2754-1169/2025.BL29505

public markets or entire sectors; they rarely test the micro link between micro entrepreneurs' public social media traces and actual loan defaults. Meanwhile, big data credit reporting still faces legal, managerial, and cybersecurity hazards and requires a shared credit information exchange platform, wider digital coverage, and improved effectiveness of credit information application and regulation [4, 5].

This paper will pinpoint two core weaknesses of legacy credit bureaus: data poverty and information islands. Sort out three AI-plus-big-data applications: chain diffusion pattern recognition, multi-dimensional profiling, and real-time default prediction, and explore the practical route to safer AI-driven credit reporting: data compliance, privacy-preserving computation, and model explainability.

2. Research methods

This study mainly adopts two research methods: literature research method and comparative analysis method, and ensures the scientificity and depth of the research through systematic research design.

In terms of literature research, the research focus on the limitations of traditional credit reporting systems and the application of " AI + big data" in credit reporting. During the research process, the research identified several key search terms such as " traditional credit investigation", " AI credit investigation", and " big data credit evaluation", and used these keywords to lock in relevant literature. The literature sources include CNKI, Google Scholar and other professional databases. At the same time, it also collects public credit report samples, financial statements of enterprises in recent years and relevant reports of financial media to ensure the diversity and comprehensiveness of data sources. When screening the literature, the research focus on the results in the field of credit reporting in the past five years, mainly selecting core journal papers, excellent master's and doctoral papers, and excluding documents with weak correlation or low academic value.

In terms of comparative analysis, the research mainly carry out from two dimensions : one is to compare the research of scholars in different periods vertically, sort out the theoretical development process from traditional credit reporting to AI credit reporting, and understand the change of research focus; The second is to compare the views and methods of different researchers in the same period horizontally, analyze the similarities and differences between the academic circles on key issues such as credit data dimension and AI model application, and explore the reasons for these differences. Through the combination of these two methods, the researchcan not only fully grasp the existing research results, but also deeply analyze the internal logical relationship.

3. Structural defects of the traditional credit system

As inclusive finance advances, the built-in flaws of the legacy credit default have become glaring. They mainly reflect in three areas:

3.1. One-dimensional data

Most scholars regard a narrow data scope as the cardinal weakness, which locks "thin-file" consumers, those without any loan history, out of the formal financial system [5].

Traditional bureaus rely almost exclusively on bank-credit records, lacking unstructured signals such as e-commerce turnover or social media activity. However, recent work shows that weak

DOI: 10.54254/2754-1169/2025.BL29505

variables like GMV or social-platform engagement add significant explanatory power to default models, yet they are still absent from the central-bank file.

As figure 1 illustrates, the conventional system contains 525 variables, among which are 500 structured and only 25 unstructured, covering 37.6 % of individuals and lower to 28.4 % of firms. An AI-enriched multi-dimensional bureau expands the pool to 3,150 variables, among which are 800 structured, 2,350 unstructured, lifting coverage to 70.0 % of individuals and 76.8 % of companies.

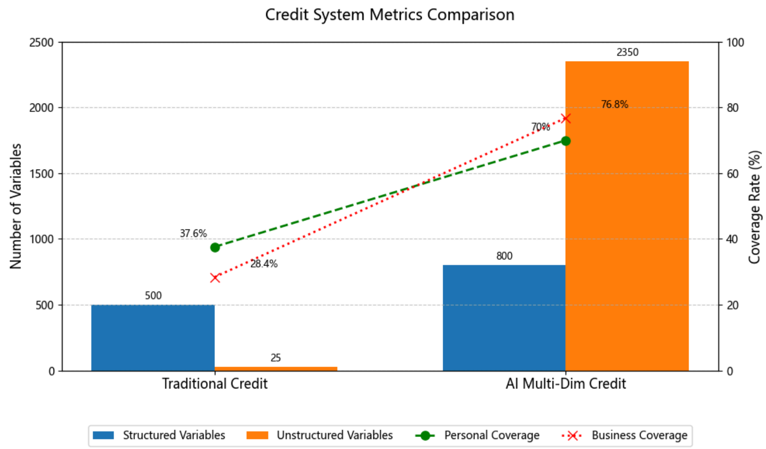

Figure 1. Comparative indicator profile of credit-reporting systems

3.2. Information update lags behind

The central-bank credit file is refreshed regularly, monthly or quarterly, and the typical report lags reality by 45–60 days. It is unable to reflect the borrower's real-time credit status, leading to a weakening of risk signals [4]. Insufficient timeliness is attributed to batch file submission rules and obsolete tech stacks [5]. In the era of the digital economy, this lag makes it difficult for financial institutions to identify potential risks on time.

3.3. Severe information islands

Traditional credit data are scattered across commercial banks, consumer-finance companies, and other institutions. The lack of a unified sharing mechanism results in banks, tax authorities, and the judicial system not being interconnected, creating a “data chimney” [5]. The cost of collecting credit information is high, security is difficult to ensure, and no unified standards have been established [6]. This patchwork restricts the comprehensiveness and accuracy of credit evaluation. In summary, the traditional bureau’s shortcomings in data breadth, timeliness, and shareability leave ample room for AI and big-data technologies to step in.

4. AI-driven credit reporting: risks and governance challenges

While AI plus big data has markedly raised the efficiency of credit systems, its actual application process is still accompanied by multiple risks and challenges.

4.1. Legal-compliance risk

Currently, although the Personal Information Protection Law provides regulations on data usage, such as Article 13 requiring the "minimum necessity" principles to set a compliance boundary for

DOI: 10.54254/2754-1169/2025.BL29505

unstructured data. AI credit actual application, such as social media and e-commerce orders, is ambiguous, leading to disputes over "over-collection" and making it easy to trigger data leakage incidents [4].

4.2. AI privacy leakage risk

Big data credit reporting requires multi-source data integration as support, which significantly increases the possibility of data exposure. If the identification and anonymization are skipped during ingestion or model training, it may lead to exposure of user privacy. Therefore, the concept of privacy protection first must be embedded in the technical architecture design, actively adopting various privacy computing technologies to strictly control the risk of privacy leakage while ensuring data utility.

4.3. Insufficient model interpretability

Although deep learning models such as random forests and neural networks have high prediction accuracy, their black-box nature makes it almost impossible for Financial institutions to explain the reason for the rejection of an applicant. It makes its decision-making process difficult to fully explain [5]. Financial institutions may face regulatory penalties for being unable to explain the reasons for loan refusal, which also affects user trust and the acceptability of the model.

5. The application and development of AI and big data technology in the field of credit information

5.1. Construction of association networks: breaking data barriers

In the construction of related networks, the credit data of small and micro enterprises scattered in financial institutions and e-commerce platforms in traditional credit reporting is fragmented. Through the chain diffusion algorithm, AI combines graph neural network and knowledge graph to integrate fragmented information such as corporate social relations, e-commerce transaction links and public records to build a panoramic credit network. The network can accurately identify the risk transmission chain. For example, when the core supplier defaults, it can warn the performance risk of the affiliated enterprise in real time, promote the credit evaluation from “ individual analysis” to “ associated network evaluation”, and effectively break the data barrier.

5.2. Full-dimensional credit portrait : breaking through data limitations

The construction of full-dimensional credit portrait solves the pain point that traditional credit reporting is difficult to cover “ credit white households”. AI and big data transform unstructured information into quantifiable credit indicators, using natural language processing to analyze corporate social media data and extract soft dimensions such as business stability ; modeling e-commerce transaction data based on machine learning to quantify business capabilities ; integrate tax, social security and other public data to improve compliance assessment. The neural network model of Khashman and the random forest tree algorithm of Ghatasheh all confirm the accuracy advantage of AI model in credit risk prediction, and help credit reporting upgrade from “single financial dimension” to “comprehensive behavior portrait” [7, 8].

Proceedings of ICFTBA 2025 Symposium: Data-Driven Decision Making in Business and Economics
DOI: 10.54254/2754-1169/2025.BL29505

5.3. Dynamic risk early warning : real-time monitoring

Dynamic risk early warning realizes the transformation of credit reporting from lag response to real-time monitoring. Different from the traditional mode of credit reporting updated monthly or quarterly, with an average lag of 45-60 days, AI collects real-time data such as transaction flow and business dynamics of enterprises, and continuously monitors credit status in combination with dynamic modeling. Wang improved the GA-BP neural network to provide technical support for the dynamic risk assessment of big data in Internet financial credit reporting [9]. Chen Liuqin pointed out that the AI risk control system can dynamically adjust the parameters to predict the probability of default [10]. The research of Cheng Wengeng and Li Rui also confirmed that dynamic data can improve the accuracy of default prediction and promote risk identification to shift from post-analysis to in-process intervention [1].

At the market level, the scale of AI and big data related applications continues to expand. The global fraud detection market will reach USD 5.282 billion in 2024, and is expected to increase to USD 19.745 billion in 2030, as can be seen in figure 2. The growth rate of generative AI in the field of financial services is 26.29 %. It can help credit reporting agencies to reduce costs and increase efficiency, such as the intelligent upgrade of the full credit reporting access model, and promote the development of the credit reporting industry in an efficient and accurate direction, laying a foundation for inclusive finance to cover more white households and small and micro operators.

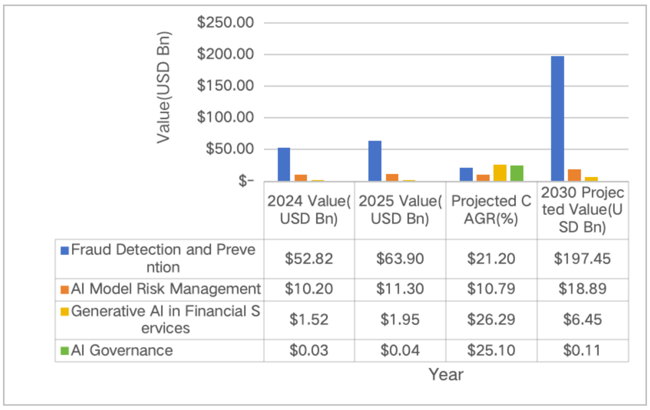

Figure 2. AI-enabled risk and compliance market size and forecast (2024-2030)

6. Conclusion

Through literature research and comparative analysis, this study explores the application value and risks of traditional credit reporting defects, AI and big data. The conclusions are as follows. Traditional credit reporting has a single data dimension (only 25 unstructured variables, personal and enterprise coverage 37.6 %, 28.4 %), information update lag (45-60 days), and serious information islands. It is difficult to adapt to the needs of inclusive finance ; aI and big data break through the traditional limitations with the help of associated network construction, full-dimensional credit portrait and dynamic risk early warning. In 2024, the global fraud detection market reached USD 5.282 billion, but it faces legal compliance, privacy disclosure and model “ black box” risks.

Based on this, some suggestions are put forward. technically, optimizing federal learning to adapt small and micro data, and strengthening the interpretability of AI model; in the system, a dual mechanism of “compliance framework + privacy calculation” is established to clarify the boundary

DOI: 10.54254/2754-1169/2025.BL29505

of data collection; ecologically, accelerate the construction of credit information sharing platform, use AI to enhance regulatory capacity and verify new technologies through regulatory sandboxes.

There are limitations in this study. focusing on macro and industry analysis, not empirically testing the micro-association between social media information of small and micro operators and loan defaults, and not in-depth comparing the scene suitability of different AI models. The micro-applicability of the conclusion and the pertinence of the model need to be improved.

Looking forward to the future, it is necessary to promote the transformation of the credit reporting industry with technological innovation, system improvement, and ecological construction: continuously explore the performance of the federal learning optimization model, improve the data compliance and privacy protection system, and accelerate the landing of the credit information sharing platform, and finally form a virtuous circle of data sharing expansion coverage, technological innovation to improve accuracy, and regulatory protection to promote security, so as to promote the development of credit reporting in the direction of precision, inclusiveness, and security, and serve the digital economy and inclusive finance.

Authors contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

[1] Cheng, W. G., & Li, R. (2022). Research on the construction of credit information system for financial leasing industry based on big data risk control: A case study of the grain and oil processing industry. Credit Reference, 40(6), 64–70.
[2] Doğuş Emin, A. A. E. (2025). The role of AI in credit risk assessment: Evidence from OECD and BRICS via system GMM and random forest. Finance Research Letters, 81, 107499. https: //doi.org/10.1016/j.frl.2025.107499
[3] Bao, W.Q., Xu, K., & Leng, Q. (2024). Research on the financial credit risk management model of real estate supply chain based on GA-SVM algorithm: A comprehensive evaluation of AI model and traditional model. Procedia Computer Science, 243, 900–909. https: //doi.org/10.1016/j.procs.2024.09.108
[4] Wang, H.B. (2021). Credit risk management of consumer finance based on big data. Mobile Information Systems.
[5] Zhou, L., Gu, R. P., Xing, X., & Wu, H.Y. (2022). A review and prospect of frontier research on big data credit reporting. Credit Reference, (1), 36–44.
[6] Zhang, C. (2022). Research on credit data sharing mechanism based on blockchain. Jilin Financial Research, (01), 60–64.
[7] Khashman, A. (2010). Neural networks for credit risk evaluation: Investigation of different neural models and learning schemes. Expert Systems with Applications, 37(9), 6233-6239.
[8] Ghatasheh, N. (2014). Business analytics using random forest trees for credit risk prediction: a comparison study. International Journal of Advanced Science and Technology, 72(2014), 19-30.
[9] Wang, F. P. (2018). Research on Application of Big Data in Internet Financial Credit Investigation Based on Improved GA-BP Neural Network. Complexity, 2018(1), 7616537.
[10] Chen, L. Q. (2025). Credit reporting innovation from the perspective of new-quality productive forces: Exploring the path of finance-technology integration. Credit Reference, (4), 21–29.