Journal of Frontier in Economic and Management Research

DOI URL:https://doi.org/10.63944/2q3JFEMR

Vol.1 No.1

Transformative Impacts of Big Data Technologies on the Credit Reporting Industry: Drivers, Challenges, and Future Trajectories

Zhiming Song[1*], Huilin Mo[1]

[1] School of Economics, Jiangsu Normal University Kewen College, Xuzhou, Jiangsu, China | sarahsonhengmincamille@gmail.com

[*]Corresponding Author: Zhiming Song | sarahsonhengmincamille@gmail.com

Abstract: Amid the rapid evolution of the digital economy, big data technologies are reshaping the foundations of traditional credit reporting by expanding data sources, refining modeling methods, and enhancing risk response capacity. From the integrated perspective of the "technology–institution–ethics" triad, this paper systematically reviews 33 studies published between 2012 and 2025, supplemented by representative case analyses. The review follows PRISMA 2020 guidelines, covering both international literature and China-specific practices. The analysis shows that while big data enables more dynamic, precise, and intelligent credit evaluation, it also generates systemic risks, including privacy infringement, algorithmic bias, model opacity, and regulatory lag. To address these dilemmas, a comprehensive governance framework is proposed that combines explainable artificial intelligence, privacy-preserving computation, cross-sector regulatory coordination, and ethical algorithmic norms. The study acknowledges its limitations as a review-based work—particularly in terms of proprietary data accessibility, interpretability of complex models, and empirical cross-platform validation—and suggests future research directions involving real-world experimentation, interpretable deep models, and multi-institutional governance mechanisms. Overall, this research aims to provide theoretical foundations and policy insights for building an open, transparent, and sustainable digital credit ecosystem.

Keywords: Big data credit reporting; explainable AI; algorithmic governance; regulatory coordination; digital credit ethics

1. Introduction

In modern financial systems, the credit mechanism serves as a cornerstone for both resource allocation and risk management. An efficient and transparent credit reporting system not only enables financial institutions to assess borrowers' credit risk, reduce default probabilities, and minimize non-performing assets, but also plays a crucial role in advancing financial inclusion and enhancing the operational efficiency of financial markets. Traditional credit reporting systems rely primarily on structured financial data—such as repayment records, balance sheets, and historical loan behaviors—which are processed through conventional scoring models to assess creditworthiness. While this model has established a relatively robust mechanism for credit evaluation over time, it has increasingly revealed limitations such as data homogeneity, delayed updates, and narrow dimensions of risk identification. These shortcomings render it inadequate for addressing the complex demands of dynamic credit assessment in the digital economy.

With the rapid advancement of information technologies—particularly the widespread adoption of big data—credit reporting systems are undergoing profound transformations. Characterized by the “3Vs” of big data (Volume, Variety, and Velocity), big data offers a novel foundation for data acquisition and model construction in credit evaluation. By integrating unstructured and semi-structured data from online behavior, social networks, e-commerce transactions, and geolocation information, big data-based credit reporting enables the creation of more comprehensive and multidimensional credit profiles, thereby enhancing both the coverage and precision of credit assessments (Nahar et al., 2024). For instance, fintech firms such as ZestFinance and Ant Group’s Sesame Credit in China have applied AI-driven big data models across loan approval and risk control scenarios, achieving greater efficiency in risk identification and dynamic evaluation (Sadula, 2023).

However, alongside the restructuring of credit systems, big data credit reporting also introduces a series of new risks and challenges. On the one hand, issues such as inconsistent data quality, redundant or fabricated information, and the opacity of so-called “black-box models” may undermine transparency and reliability, potentially leading to algorithmic discrimination and systemic bias (Chang & Li, 2018). On the other hand, the cross-platform circulation of data has intensified concerns over privacy and information security, raising difficult regulatory questions about how to balance data utilization with the protection of individual privacy (Liu & Hou, 2021). Moreover, the current credit reporting market faces structural barriers such as data fragmentation, entrenched “information silos,” and the absence of unified industry standards, all of which hinder the integration and sustainable development of big data credit systems.

Against this backdrop, this paper presents a systematic review of the transformation pathways and institutional responses in the credit reporting industry driven by big data technologies. The review is explicitly grounded in 33 academic and professional publications spanning 2012–2025, identified through major international databases and authoritative Chinese sources, and follows PRISMA 2020 guidelines to ensure transparency in identification, screening, and synthesis. The study addresses three core questions: (1) How does big data technology drive the functional restructuring and value enhancement of credit systems? (2) What technical, ethical, and governance challenges are encountered in its practical application? (3) What are the future trajectories of big data credit reporting, and can it replace or effectively integrate with traditional credit mechanisms?

To answer these questions, the study adopts a structured framework—drivers, challenges, and future trajectories—through which it synthesizes international research findings and representative case studies. The aim of this research is twofold: theoretically, to deepen the understanding of the interaction between credit institutions and technological innovation, and to enrich paradigms of credit reporting evolution; and practically, to provide references for policymakers, financial institutions, and credit service providers, thereby supporting the development of a more efficient, equitable, intelligent, and trustworthy modern credit reporting system.

2. Research Methods

2.1. Study design

This study employs a systematic literature review (SLR) supplemented by targeted case analyses. The SLR provides a structured and reproducible foundation for synthesizing evidence on how big data technologies transform credit reporting, while the case analyses highlight specific implementation pathways and governance practices. The review process follows the PRISMA 2020 guidelines to ensure transparency in study identification, screening, eligibility assessment, and inclusion.

2.2. Data sources and search strategy

The review draws on 33 publications retrieved from multiple academic and professional databases, including Web of Science, Scopus, IEEE Xplore, SSRN, ProQuest, and CNKI, complemented by relevant books, theses, and authoritative industry reports. The search window spanned 2012–2025, covering both early conceptual debates (e.g., algorithmic governance and privacy concerns) and recent empirical advances (e.g., federated learning and blockchain-based credit scoring).

Representative search queries combined the following terms:

("big data" OR "data-driven" OR "alternative data" OR "federated learning" OR "machine learning")

AND

("credit reporting" OR "credit scoring" OR "credit risk" OR "digital credit" OR "credit reference")

Both English and Chinese sources were included to capture global perspectives as well as China-specific developments, given China's leading role in data-driven credit systems.

2.3. Inclusion and exclusion criteria

To ensure rigor and relevance, the review applied clear inclusion and exclusion criteria. Eligible sources comprised peer-reviewed journal articles, conference proceedings, theses, book chapters, and authoritative industry or organizational reports published between 2012 and 2025. Studies were required to focus on big data applications in credit reporting, credit scoring, or credit risk management, and to address at least one dimension of the topic—whether technological, ethical, regulatory, or institutional. By contrast, materials such as news items, non-academic commentaries, or documents lacking methodological or conceptual depth were excluded. Duplicate records or secondary sources without original contributions were also removed, as were studies unrelated to credit reporting (for example, those examining general big data analytics outside the financial domain).

2.4. Screening and selection process

All identified records underwent a two-stage manual screening. Titles and abstracts were first reviewed to exclude irrelevant works, after which the full texts of potentially eligible studies were assessed against the inclusion and exclusion criteria. This process resulted in a final sample of 33 publications, which together capture the breadth of existing research. The included studies reflect diverse methodological orientations: some present empirical modeling, others provide theoretical and ethical analyses, several examine China's large-scale practices through case studies, and others explore technical innovations such as federated learning.

2.5. Data extraction and synthesis

For each study we recorded bibliographic details (author, year, country), research objectives, methodology, data sources, main findings, and policy implications. The synthesis employed a three-pronged approach. First, descriptive mapping was used to trace temporal publication patterns (2012–2025), thematic clusters such as drivers, challenges, and governance trajectories, and the methodological diversity across the field. Second, thematic coding grouped studies into categories of technological enablers, ethical and legal challenges, institutional governance, and future trajectories. Finally, case illustration highlighted representative real-world applications—including Ant Group's Sesame Credit, ZestFinance, and blockchain-enabled federated learning—to contextualize broader trends and governance implications.

The detailed characteristics of the 33 included studies are presented in Table 1, which summarizes their methodologies, principal findings, and policy implications. The table demonstrates the diversity of approaches—ranging from conceptual analyses and legal-normative studies to empirical modeling and technical innovations—and highlights the interplay between technological development, ethical considerations, and regulatory responses. This structured summary provides the empirical foundation for the thematic synthesis that follows in the subsequent sections.

Table 1. Summary of included studies on big data in credit reporting (2012–2025)

Author(s), Year	Methodology	Main Findings	Policy / Governance Implications
Nahar et al., 2024	Systematic review	Synthesizes transformative practices of big data in credit risk management	Need for integrated governance and future-oriented frameworks
Sadula, 2023	Empirical (data integration)	Combines SEC datasets with Altman Z'-Score model	Enhances transparency; regulatory oversight of alternative analytics
Chang & Li, 2018	Conceptual / ethical analysis	Discusses ethical issues in big data credit reporting	Emphasizes privacy protection and ethical standards
Liu & Hou, 2021	Case study (China)	Examines challenges of big data-based credit reference	Highlights data quality and regulatory gaps in China
Xie, 2023	Empirical (bank applications)	Applies big data to commercial bank credit business	Encourages adoption with stronger data governance
Wang, Zhang & Li, 2020	Empirical	Analyzes social big data's impact on personal credit	Stresses risks of bias and misuse of social data
Wang, 2021	Theoretical / applied model	Proposes big data-based consumer finance risk management	Supports integration with consumer protection frameworks
Yang et al., 2018	Case study	Builds credit reporting model for SMEs in Humen	Shows potential of sector-specific big data models
Ransbotham, 2016	Applied case	Explores unstructured data in credit reporting	Promotes innovation but warns of governance needs
Gao & Xiao, 2021	Empirical model	Tests big data credit report in consumer finance	Finds improved risk control; calls for transparency
Cui, 2015	Conceptual / modeling	Proposes credit risk warning system based on big data	Advocates early-warning tools for regulators

Sun, 2021	Case study (Bairong Zhixin)	Explores big data in China's credit system	Shows importance of regulatory alignment
Liu et al., 2017	Empirical model	Uses telecom/mobile data for credit scoring	Promotes inclusion but raises privacy concerns
Shi, 2012	Case study	Describes China's national credit scoring system	Demonstrates early state-led big data credit efforts
Yiyuan et al., 2019	Empirical (SMEs)	Builds SME credit assessment model	Highlights SME-focused credit policy needs
Cheung & Chen, 2017	Legal / normative	Analyzes profiling and privacy in social credit	Urges stronger legal safeguards
Zhou & Fu, 2024	Empirical (China firm data)	Tests big data platform effects on credit mismatch	Finds efficiency gains; regulatory support needed
Cui et al., 2016	Empirical framework	Proposes SME big data credit reporting	Suggests institutional innovation
Mbah, 2024	Legal/regulatory review	Analyzes data privacy under AI	Calls for harmonized privacy frameworks
Liu, 2018	Conceptual	Critiques big data credit problems in China	Suggests countermeasures to risks
Sargeant, 2022	Normative/ethical	Examines algorithmic decision-making in credit	Calls for fairness and accountability
Shittu, 2022	Technical/empirical	Proposes AI-driven credit risk models	Shows efficiency gains; stresses need for oversight
Segun-Falade et al., 2025	Legal / conceptual	Examines AI-powered privacy risk assessments	Proposes compliance-oriented legal models
Wang & Huang, 2023	Case study (China FinTech)	Analyzes FinTech–bank partnerships	Highlights risk-sharing and regulation
Harvey, 2019	Empirical	Studies big data in credit scoring & lending	Supports inclusive lending policies
Schultz & Crawford, 2013	Legal / conceptual	Analyzes due process in big data use	Proposes safeguards against predictive harms
Feng & Chen, 2022	Technical model	Blockchain + federated learning for privacy	Advocates hybrid governance models
Mangues-Bafalluy et al., 2024	Technical / conceptual	Blockchain-based reputation in FL	Supports trustworthy data collaboration
Zheng et al., 2020	Technical model	Proposes vertical FL scorecard for credit scoring	Improves interpretability of models
Pingulkar & Pawade, 2024	Technical comparative	Compares FL architectures for credit risk	Guides scalable privacy-preserving models
Song et al., 2021	Technical / applied	Trusted telecom-finance data sharing method	Ensures compliance with China's data policy
Müller et al., 2023	Socio-technical analysis	Studies challenges of federated ML	Recommends socio-technical governance
Zavolokina et al., 2024	Conceptual / applied	Addresses intergovernmental data sharing with FL	Suggests FL for cross-border governance

2.6. Conceptual research framework

To provide a clear analytical roadmap, this study develops a conceptual research framework (Figure 1) that integrates the main dimensions identified through the systematic review. The framework begins with the drivers of big data credit reporting, such as technological innovations, regulatory demand, and market efficiency needs. These drivers lead to a set of challenges, including data privacy risks, algorithmic bias, model opacity, and institutional fragmentation. In response, a variety of governance mechanisms have been proposed and implemented, ranging from explainable artificial intelligence and privacy-preserving computation to cross-sector regulatory coordination and ethical algorithm design. Finally, the framework highlights potential future trajectories, pointing to more transparent, inclusive, and sustainable digital credit ecosystems. This framework not only structures the synthesis of findings but also serves as a reference point for subsequent discussion and policy implications.

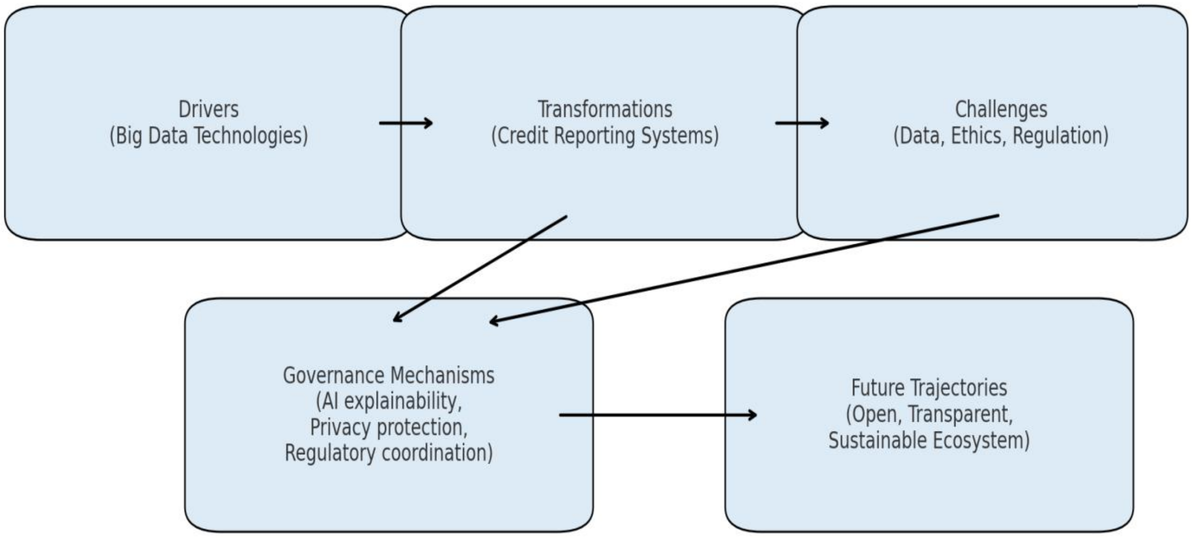

Figure 1. Conceptual research framework of big data credit reporting

3. The Core Characteristics of Big Data and Its Integration with Credit Reporting Systems

With the rapid advancement of the digital economy, big data technologies have emerged as a key driving force in reshaping the architecture and operational logic of credit reporting systems. Their impact extends beyond improvements in data processing capacity and analytical efficiency, fundamentally transforming credit assessment paradigms and stimulating corresponding institutional innovations. This section systematically examines the technical characteristics of big data and their implications for credit evaluation, while exploring the structural transformations and institutional responses they have triggered within the credit reporting domain.

3.1. From “3Vs” to “5Vs”: Paradigm Shifts in Credit Logic

The transformative influence of big data on credit systems stems primarily from its well-known “3V” characteristics—Volume, Velocity, and Variety—which have been further extended into the “5V” framework by adding Veracity and Value. These features form the technological foundation for big data applications in credit evaluation and redefine the core logic of credit identification, risk assessment, and pricing mechanisms.

Volume refers to the exponential growth in the scale and diversity of data sources. While traditional credit systems rely on structured data collected within financial institutions (e.g., credit histories and repayment records), big data credit models integrate unstructured and semi-structured data such as social media activity, e-commerce transactions, search behavior, mobile payments, and geolocation information. This significantly expands the variable space for credit modeling (Nahar et al., 2024).

Velocity enhances the timeliness of credit assessments. Unlike traditional models that rely on ex-post analysis of historical data, big data platforms enable real-time data streaming and high-frequency modeling, allowing credit systems to dynamically track changes in users’ credit status and respond promptly to emerging risks (Xie, 2023).

Variety reflects the heterogeneous integration of diverse data types and formats. Modern big data credit models incorporate not only structured financial indicators but also textual, visual, social network, and behavioral sequence data, resulting in a multi-source, multimodal analytical framework (Wang et al., 2020).

Veracity emphasizes the credibility and reliability of data. As the volume and complexity of data grow, challenges such as noise, redundancy, and falsification become more pronounced. Strengthening institutional data governance frameworks and establishing industry-wide standards are therefore essential to building robust and effective credit models (Chang & Li, 2018).

Value represents the ultimate benefit of big data in credit systems. Leveraging techniques such as machine learning, graph analysis, and behavioral modeling, large-scale data can be transformed into actionable insights, thereby improving institutions’ risk control capabilities and enabling more refined credit pricing strategies (Wang, 2021).

In sum, the 5V framework not only constitutes the technical foundation for embedding big data into credit evaluation processes but also highlights the necessity of institutional adaptation in data governance and regulatory design, driving the evolution of credit assessment from static, linear models to dynamic and intelligent paradigms.

3.2. Structural Bottlenecks and Failures in Traditional Credit Systems

Despite their foundational role in financial resource allocation, traditional credit systems face increasing structural limitations, manifesting in three primary areas:

First, information asymmetry remains difficult to resolve. Traditional systems rely heavily on financial institution-reported data, excluding behavioral data from non-financial contexts. This leads to underrepresentation of “credit-invisible” and marginalized populations, resulting in unfair assessments and potential systemic financial exclusion (Li & Yang, 2018).

Second, issues of pricing bias and model discrimination are prominent. Traditional scoring models often depend on linear variables and static indicators, overlooking critical behavioral features such as psychological stability, social network cohesion, and short-term liquidity. Some models also embed discriminatory assumptions based on gender or region, undermining fairness and inclusivity in credit systems (Ransbotham, 2016).

Third, the lack of model transparency and regulatory oversight raises ethical concerns. Traditional models offer limited interpretability, making it difficult for the public to understand scoring rationales. This fuels fears of “algorithmic black boxes” and data misuse, while regulatory mechanisms have yet to effectively penetrate the modeling and data processing stages (Chang & Li, 2018).

In conclusion, these bottlenecks highlight the systemic failures of traditional credit systems: persistent information asymmetry, embedded biases, and insufficient governance. Such failures underscore the urgency of adopting big data-driven approaches that combine technological innovation with institutional reform.

3.3. From Static to Dynamic Credit: A Paradigm Shift in Evaluation

One of the most significant innovations in big data credit reporting is the shift from static credit evaluation to dynamic credit modeling. This transformation enhances real-time responsiveness, personalization, and predictive power in credit systems.

Conventional credit scoring is often based on historical “snapshots” and fails to capture short-term fluctuations in creditworthiness. For instance, sudden unemployment or atypical behaviors may not be promptly detected. In contrast, big data-enabled dynamic modeling incorporates time-series and high-frequency behavioral data to continuously track users’ evolving risk trajectories (Gao & Xiao, 2021). Atypical patterns such as sharp declines in spending, frequent location changes, or reduced social engagement can automatically trigger warning mechanisms, prompting real-time adjustments in credit ratings and lending decisions (Cui, 2015).

This enhances the responsiveness and intelligence of risk control systems, effectively mitigating issues of delayed risk recognition and systemic inertia.

Furthermore, regulatory institutions are beginning to explore dynamic supervision mechanisms that align with real-time risk monitoring, indicating that technological innovation and institutional innovation must proceed in tandem.

3.4. Reconstruction of Credit Dimensions and the Emergence of Closed-Loop Data Logic

A deeper evolution of big data credit reporting lies in the reconstruction of credit dimensions and the establishment of closed-loop evaluation mechanisms. Credit is no longer confined to quantifying financial transactions but has expanded into a composite system encompassing social behavior, lifestyle patterns, and online reputations.

New variables—such as social stability, behavioral consistency, and payment rhythm regularity—allow models to infer internal credit intentions and fulfillment propensities from external behaviors. This supports a more nuanced, dynamic, and personalized evaluation logic (Li et al., 2020).

In practical applications, big data credit systems have gradually developed a closed-loop process encompassing risk identification, trend prediction, dynamic scoring, and automated alerts. Through feature engineering and multi-source data fusion, potential risks are identified; machine learning techniques are used for predictive analysis; multidimensional indicators are integrated to generate personalized scores; and credit ratings are dynamically updated to inform real-time credit decisions (Nahar et al., 2024).

Importantly, this closed-loop logic also pushes forward institutional innovation: regulators are encouraged to establish adaptive supervisory frameworks, financial institutions are required to improve internal compliance mechanisms, and industry standards must evolve to ensure interoperability across platforms.

This closed-loop mechanism enhances the scientific rigor, predictive accuracy, and forward-looking capacity of credit models, offering a viable solution to longstanding issues such as limited coverage, distorted scores, and delayed feedback in traditional systems. It marks a significant step toward the maturation and practical application of a data-centric and institutionally coordinated credit logic.

4. Key Impacts of Big Data Technologies on the Credit Reporting Industry

As big data technologies become deeply embedded in the financial sector, the credit reporting industry is undergoing a systemic and multidimensional transformation—ranging from data architecture and modeling logic to risk management frameworks and service ecosystems. In contrast to traditional credit systems, which rely primarily on static models built from limited financial transaction data, big data enables the reconstruction of information sources, the algorithmic upgrading of modeling approaches, the dynamic reinforcement of risk control, and the restructuring of business models. These changes have not only improved the accuracy and timeliness of credit evaluations but also generated profound impacts on financial institutions, regulators, enterprises, and individual consumers. This section examines these impacts across four dimensions.

4.1. Diversification of Data Sources: Reshaping the Structure of Credit Information

For decades, traditional credit systems relied predominantly on structured financial data—such as repayment histories, credit card usage, and loan records. While authoritative, such data sources suffer from limited coverage, lagged updates, and insufficient explanatory depth, particularly when assessing "credit-invisible" groups such as rural migrants, small and micro enterprises, or young consumers. This has led to persistent "credit blank zones."

The integration of big data has substantially expanded the informational boundaries of credit evaluation, encompassing behavioral, transactional, and contextual data drawn from domains such as e-commerce, mobility, communication, and geolocation (Sun, 2021). These new data streams, often referred to as "soft information," complement or even substitute traditional financial data in risk identification (Liu et al., 2017). For example, systems such as Sesame Credit and Tencent Credit now integrate shopping frequency, social network density, contract performance, and mobility stability into their models. This evolution represents a shift from "financial credit" to a fusion of "behavioral credit" and "contextual credit" (Wang et al., 2020).

Importantly, stakeholder impacts are also evident: financial institutions can reach underbanked customers, small businesses gain broader financing channels, and consumers receive differentiated credit services. The trend of multi-source data integration is therefore restructuring the informational foundation of credit reporting and expanding its inclusiveness.

4.2. Intelligent Modeling: From Rule-Based Logic to Machine Learning

The added value of big data lies not only in richer inputs but also in enabling a paradigm shift in credit modeling—from rule-based scoring to machine-learning-driven intelligence. Traditional models, typically logistic regression, excel in interpretability but falter when handling nonlinear patterns or high-dimensional variables.

Machine learning techniques—such as Random Forest, XGBoost, Support Vector Machines, and deep neural networks—allow credit agencies to process high-dimensional user, network, and behavioral data with greater predictive accuracy and adaptability (Shi, 2012). National projects such as the "China Score" have leveraged knowledge discovery and data mining (KDD) to create personalized and behavior-driven scoring frameworks with enhanced robustness.

For small and micro enterprises, where traditional balance-sheet-based assessments are prone to manipulation and time delays, big data models integrate supply chain activities, online reputation, and social networks, thereby enabling real-time evaluation of operational health and repayment capacity (Liu et al., 2019). The modeling logic of credit systems has thus evolved into a self-learning, data-driven framework that is dynamic, adaptive, and context-sensitive.

4.3. Strengthened Risk Management: From Static Assessment to Dynamic Prevention

Big data has transformed credit risk management from pre-loan assessment into full-lifecycle monitoring, covering pre-loan screening, in-loan surveillance, and post-loan early-warning systems. By integrating multi-source behavioral and transactional data with high-frequency features, credit systems can now detect financial fraud, credit deterioration, and anomalous patterns in real time.

Specifically, the analysis of location data, device fingerprints, login patterns, and social interaction sequences enhances the ability to identify risks such as identity theft, account hijacking, and cross-platform fraud (Chen & Cheung, 2017). For instance, Sesame Credit's tracking of contractual behavior produces long-term trend maps that effectively flag potential defaults (Li et al., 2020).

Through behavioral sequence analysis and graph-based algorithms, credit systems are shifting from reactive, static models toward proactive prevention and real-time correction, enabling the construction of self-adaptive risk control ecosystems. For stakeholders, this reduces systemic risk for regulators, enhances portfolio stability for lenders, and strengthens consumer protection.

4.4. Transformation of Business Models: Platformization, Servitization, and Openness

The influence of big data extends beyond data and modeling to redefine the organizational and commercial structure of credit services. Traditional systems were centralized and monopolistic, with a handful of institutions controlling credit data and scoring, which limited market diversity and accessibility.

Currently, the industry is shifting toward platform-based, service-oriented, and open ecosystems. The rise of "Credit-as-a-Service" (CaaS) allows agencies to deliver scoring modules, risk alerts, and credit tags to third parties via APIs, enabling modular and scenario-specific applications (Fu & Zhou, 2024). This has created embedded credit services across e-commerce, mobility, housing, and education, significantly expanding usage contexts.

Meanwhile, cross-platform data integration and government-supported information sharing are dismantling “information silos.” For example, local enterprise credit platforms now aggregate tax, registration, and online transaction data to provide unified scoring models, facilitating collaboration with banks, insurers, and guarantors (Sun et al., 2016). Platformization thus enhances service scalability, industry inclusiveness, and social trust in credit markets.

In sum, big data is reshaping credit reporting across four interlinked dimensions: diversified data sources, intelligent modeling, dynamic risk management, and platform-based service ecosystems. These transformations extend beyond technical upgrades, generating new opportunities and challenges for regulators, financial institutions, enterprises, and consumers alike. The credit reporting industry is therefore entering a new phase defined by openness, intelligence, inclusiveness, and stakeholder co-evolution.

5. Challenges and Risks in Big Data-Based Credit Reporting

Despite the transformative potential of big data in enhancing the intelligence and inclusiveness of credit reporting systems, its widespread application has also introduced new layers of institutional, technical, and ethical risks. Under conditions of high-frequency data collection, algorithmic modeling, and cross-domain embedding of credit evaluations, the operational mechanisms of credit systems have grown increasingly complex. The resulting risks now extend beyond technical performance to deeper concerns of governance, fairness, and social trust. This section examines five key dimensions of these challenges.

5.1. Data Privacy and Information Security

Big data credit platforms process vast volumes of highly sensitive information—including financial transactions, consumption habits, geolocation, and social networks. While such integration enhances precision, it also amplifies the risk of privacy breaches and unauthorized data flows.

Although regulatory frameworks such as the Personal Information Protection Law and the Data Security Law in China provide institutional boundaries emphasizing data minimization, informed consent, and lawful processing (Mbah, 2024), violations remain frequent. Common practices include default authorization, bundled consent, and opaque secondary uses of data (Liu, 2018). These weaken users’ rights to know and to control their data, eroding public trust and institutional legitimacy. For stakeholders, individuals face privacy exposure, enterprises bear compliance risks, and regulators struggle with enforcement gaps.

5.2. Algorithmic Discrimination and the “Black Box” Problem

The shift to machine learning and deep learning has heightened opacity in credit scoring. While traditional statistical models were already not fully transparent to lay users, they were at least rule-based and interpretable to experts. By contrast, big data models—especially neural networks—generate outcomes through complex, non-linear processes that resist explanation, exacerbating the "black box" problem (Macmillan, 2019).

Moreover, training data often contain historical biases or proxy variables such as occupation, residence, or purchasing patterns. Even if sensitive variables like gender or ethnicity are excluded, proxies may still reproduce discriminatory effects (Sargeant, 2022). The result is the risk of systematically disadvantaging certain groups, thereby reinforcing inequality under the guise of neutrality. Addressing these challenges requires models with greater interpretability, fairness, and accountability, alongside mechanisms for users to appeal and contest scoring outcomes.

5.3. Data Quality and Model Distortion

Unlike traditional systems built on structured financial data, big data credit platforms rely on heterogeneous, dynamic, and unstructured information, such as online reviews, behavioral traces, and textual tags. Without standardized protocols for data cleaning and semantic alignment, issues such as label inconsistencies, ambiguous meanings, and sampling biases frequently arise (Shittu, 2022).

In addition, data sources are often contaminated with fabricated behaviors, fake reviews, or manipulated transactions. If such noise infiltrates training datasets, it can distort model predictions and erode scoring credibility (Chintoh et al., 2025). A further problem is the lack of transparent appeal and correction mechanisms: most platforms provide limited recourse for individuals to contest erroneous scores. This not only damages consumer rights but also exacerbates financial exclusion for vulnerable groups.

5.4. Regulatory Lag and Grey-Zone Practices

The pace of technological evolution has consistently outstripped regulatory capacity. In the absence of mature and enforceable norms, some credit agencies exploit grey areas by deploying unauthorized web crawlers, outsourcing data harvesting, or using illegal APIs to expand data pools. In extreme cases, illicit data markets have emerged, undermining both user rights and the credibility of credit infrastructures (Huang & Wang, 2023).

To improve model performance, platforms sometimes construct panoramic data profiles without adequate user consent. Although such practices yield short-term technical gains, they compromise contractual legitimacy and social acceptance. A responsibility-chain governance model, which clearly delineates rights and duties across collection, processing, and application phases, is urgently needed to ensure transparency, accountability, and user protection (Harvey, 2019).

5.5. Ethical Dilemmas and the Erosion of Trust

Big data-based credit systems are not merely technical tools but instruments of social governance. By embedding credit scores into access to loans, housing, transportation, and education, these systems increasingly shape life chances and social opportunities.

While such mechanisms may incentivize compliance, they also risk creating "credit penalties" that extend beyond financial contexts, restricting mobility, employment, or access to services. The downside is that such penalties can entrench social stratification, reduce upward mobility, and stigmatize already disadvantaged groups, thereby amplifying inequality.

At the same time, the construction of personalized credit profiles by commercial platforms—often without full user awareness or consent—challenges individual autonomy and freedom of choice. As Crawford and Schultz (2013) warn, without fairness, interpretability, and accountability, data-driven credit systems risk degenerating into mechanisms of surveillance and behavioral control. Establishing an ethical governance framework grounded in social justice has therefore become imperative to maintain both legitimacy and public trust.

5.6. Interconnections Among the Challenges

These five challenges are not isolated but mutually reinforcing. Privacy risks feed into regulatory gaps, as insufficient oversight enables excessive data harvesting; data quality issues exacerbate algorithmic discrimination, producing biased outcomes; and ethical dilemmas emerge as the cumulative effect of these technical and institutional weaknesses. Together, they form a systemic risk structure: without coordinated governance, big data credit systems may enhance efficiency at the expense of fairness, accountability, and trust. Addressing these risks thus requires an integrated approach that combines legal safeguards, technical standards, ethical principles, and stakeholder participation.

6. Future Development Trends and Policy Recommendations

The deep integration of big data technologies into the credit reporting sector is accelerating the shift toward more intelligent, decentralized, and platform-based architectures. While these transformations significantly enhance efficiency and inclusiveness, they also generate new institutional, regulatory, and ethical challenges. To ensure the long-term sustainability, legitimacy, and fairness of credit systems, forward-looking strategies are required across technological, institutional, regulatory, and ecosystem dimensions.

6.1. Technological Integration: Addressing Data Reliability, Transparency, and Collaboration

The future trajectory of credit reporting systems will likely hinge on the convergence of artificial intelligence and blockchain technologies. This convergence offers a pathway to resolve three longstanding challenges: data reliability, model transparency, and cross-institutional collaboration.

AI, with its strong capabilities in nonlinear modeling and high-dimensional feature extraction, provides advantages in behavioral identification, risk prediction, and personalized credit pricing. At the same time, blockchain furnishes a decentralized, tamper-proof infrastructure for secure data sharing across institutions (Feng & Chen, 2022). A particularly promising development is blockchain-enabled federated learning, which allows collaborative modeling while safeguarding data privacy. For instance, Javed et al. (2024) demonstrated the successful deployment of heterogeneous federated learning in a state-owned financial group, achieving enhanced prediction accuracy while maintaining data security. Such advances highlight the importance of continued investment in privacy-preserving and interoperable infrastructures to break down "data silos" and expand the technological frontiers of intelligent credit systems.

6.2. Institutional Recommendations: Enhancing Interpretability and Privacy Protection

As AI-based models become embedded in core financial infrastructures, traditional institutional arrangements struggle to meet rising demands for fairness, accountability, and compliance. Two policy directions stand out:

First, institutionalize transparent modeling practices. XAI techniques should be embedded into credit scoring to ensure that model logic, variable selection, and decision outcomes are comprehensible to users, regulators, and auditors (Pingulkar & Pawade, 2024). This is essential for mitigating "black box" concerns, strengthening compliance, and fostering public trust.

Second, build privacy-preserving collaboration mechanisms. With the maturation of differential privacy, homomorphic encryption, and federated learning, cross-institutional cooperation can occur without exposing raw data (Albany & Khediri, 2023). A notable case is the Vertical Federated Learning (VFL) system co-developed by China Telecom and a state-owned bank, which enables joint risk modeling while ensuring local data storage (Liang et al., 2021). Scaling such frameworks could help balance data utility with legal compliance, thereby reconciling the tension between innovation and regulation.

6.3. Regulatory Coordination: Building Multi-Stakeholder Governance

The risks inherent in big data credit reporting cut across multiple domains—ranging from data collection and algorithmic fairness to consumer privacy and systemic stability. Existing single-agency regulatory frameworks are insufficient to address such systemic challenges. A multi-stakeholder governance model is therefore needed, involving financial regulators, data protection authorities, technology firms, and consumer organizations (Müller et al., 2023).

Financial regulators should focus on ensuring credit model robustness, minimizing false positives, and preventing financial exclusion. Data protection agencies must strengthen oversight of consent mechanisms, ensure legal cross-border data transfers, and promote standardized privacy protocols. Technology providers should assume responsibility for algorithmic transparency, third-party audits, and ethical AI adoption—reflecting the principle that technological power entails technological accountability.

On the regulatory innovation front, the Regulatory Sandbox for Federated Credit Modeling is emerging as an effective mechanism. It creates a controlled environment where new models and blockchain infrastructures can be tested before full-scale deployment (Sprenkamp et al., 2024). This experimental governance model reduces systemic risks while fostering innovation.

6.4. The Emerging Credit Ecosystem: A Multidimensional Framework

To provide a systematic understanding of potential evolutionary pathways, we outline an ecosystem framework that links technological foundations, institutional practices, and governance logic. This multidimensional view highlights that credit reporting is no longer a single-industry function but an interconnected socio-technical system.

Table 2. Future Ecosystem Map of Big Data Credit Reporting

Component	**Key Features**
Technological Base	Federated Learning, Blockchain, Differential Privacy, Homomorphic Encryption
Data Sources	Banks, E-commerce Platforms, Social Networks, Mobile Operators, Public Agencies
Model Architecture	Explainable Scoring Models + Deep Neural Networks + Graph-Based Algorithms
Business Model	Credit-as-a-Service (CaaS), enabling cross-platform deployment and API access
Governance Mechanism	Multi-agency Regulatory Coordination + Internal Corporate Audits + Ethical-Legal Embedding

Source: Compiled by the authors based on literature review

The proposed ecosystem framework illustrates how emerging big data credit systems integrate technological, institutional, and ethical dimensions to support more comprehensive and trustworthy credit evaluation. Technological advances such as federated learning, blockchain, differential privacy, and homomorphic encryption provide the computational foundation for secure, interpretable, and collaborative credit assessment. Simultaneously, the scope of data sources has expanded beyond traditional financial records to include e-commerce transactions, social media activity, telecom usage, and public agency records, offering richer contextual insights for credit evaluation.

The integration of explainable AI models with deep learning and graph-based algorithms enables the dynamic modeling of complex credit behaviors while maintaining transparency and interpretability for decision-makers. Credit-as-a-Service (CaaS) models further facilitate scalable and modular deployment across platforms, enhancing innovation and accessibility. Finally, governance mechanisms that combine multi-agency regulatory coordination, internal corporate audits, and ethical-legal embedding ensure that credit systems operate transparently, comply with regulations, and uphold social trust. Together, these elements depict a cohesive and multidimensional view of the future credit ecosystem, emphasizing the interdependence of technology, institutions, and ethics in shaping sustainable and inclusive credit infrastructures.

7. Conclusion

This study systematically examines the transformative role of big data technologies in reshaping the credit reporting industry, based on a comprehensive review of literature, industry reports, and illustrative case analyses published between 2015 and 2024 from both domestic and international sources. By explicitly situating the review within this temporal and methodological scope, the study provides a structured lens to evaluate technological evolution, practical implications, institutional challenges, and future trajectories in credit reporting.

7.1. Paradigm Reconstruction of Credit Reporting under Big Data

The research adopts a structured review methodology, integrating qualitative analysis of industry cases and synthesis of scholarly findings to identify patterns, challenges, and innovation pathways. From this methodological lens, the evolution of big data—from the initial "3Vs" (Volume, Velocity, Variety) to the more comprehensive "5Vs" (adding Veracity and Value)—emerges as a defining factor reshaping the foundations of credit reporting. The scope of data has expanded beyond structured financial records to heterogeneous sources such as consumer transactions, social media footprints, and geolocation information, enhancing the timeliness, comprehensiveness, and contextual depth of credit evaluations. Simultaneously, machine learning–driven scoring models are gradually displacing traditional linear models, enabling adaptive handling of nonlinear, high-dimensional features. These developments mark a paradigm shift toward more refined, dynamic, and intelligent credit evaluation systems.

7.2. Systemic Challenges Underlying Technological Advancements

Despite these advances, the deployment of big data technologies introduces systemic risks. Key concerns include insufficient protection of personal privacy and rising compliance pressures; opacity and potential bias in algorithmic scoring; challenges in verifying the authenticity and semantic accuracy of heterogeneous data sources; and persistence of "gray-market" operations in the absence of updated institutional safeguards. Collectively, these issues highlight the tension between innovation and systemic risk in contemporary credit reporting systems.

7.3. The Imperative of a "Technology–Institution–Ethics" Governance Triad

The sustainable development of big data credit systems requires more than isolated technological progress or fragmented institutional reforms. Instead, a coordinated governance triad is needed across the following dimensions:

Technological dimension: Advance the adoption of XAI, differential privacy, federated learning, and homomorphic encryption to improve transparency, accountability, and user engagement. Institutional dimension: Strengthen cross-sectoral coordination among financial regulators, data protection authorities, and platform operators. Innovative mechanisms such as regulatory sandboxes can be leveraged to balance compliance requirements with technological experimentation. Ethical dimension: Reinforce fairness, grievance redress mechanisms, and algorithmic accountability to safeguard individual rights, prevent technological misuse, and ensure that credit reporting technologies align with the public interest.

This integrated framework highlights that technological sophistication must be matched by institutional adaptability and ethical responsibility.

7.4. Research Limitations and Future Outlook

As a review-based study, this paper draws primarily on existing literature, industry reports, and illustrative case analyses published between 2015 and 2024 in both domestic and international sources. While this approach provides a broad overview, it also entails several limitations:

Limited data accessibility: Proprietary restrictions surrounding credit data hinder the empirical validation of algorithmic performance and systemic risks. Immaturity of interpretability techniques: Although XAI has attracted significant scholarly and policy attention, its practical application to complex deep learning models remains nascent and under-tested. Insufficient empirical evidence on cross-platform collaboration: While federated learning and blockchain solutions have been piloted in selected contexts, their scalability and governance implications across multi-institutional and cross-regional settings require systematic evaluation.

Future research could address these gaps by:

Designing empirical studies using real-world credit datasets to assess the effectiveness and risks of big data models. Exploring the suitability of explainable deep learning models for high-stakes credit decision-making. Developing and empirically testing cross-departmental governance frameworks that integrate legal compliance, technological safeguards, and ethical principles across diverse markets.

In summary, big data–driven credit reporting is poised to become a cornerstone of future credit governance. Yet its progress hinges not only on technological breakthroughs but also on the integration of institutional coordination and ethical reflection. Only through the organic fusion of the technology–institution–ethics triad can credit systems evolve toward greater efficiency, transparency, fairness, and sustainability.

Acknowledgements

The authors express their gratitude to Jiangsu Normal University Kewen College for their strong support.

This research was funded by *2025 College Student Innovation and Entrepreneurship Training Program Project* (No.: 202513988003)

References

[1] Nahar, J., Alauddin, M., Rozony, F., & Rahaman, M. (2024). Big data in credit risk management: A systematic review of transformative practices and future directions. International Journal of Management Information Systems and Data Science. https://doi.org/10.62304/ijmisds.v1i04.196

[2] Sadula, S. (2023). Integrating big data analytics with U.S. SEC financial statement datasets and the critical examination of the Altman Z'-Score model. Technological University Dublin. https://doi.org/10.21427/KJ3E-FE83

[3] Chang, V., & Li, J. (2018). A discussion paper on the grey area: The ethical problems related to big data credit reporting. In Proceedings (pp. 348–354). https://doi.org/10.5220/0006823603480354

[4] Liu, C., & Hou, C. (2021). Challenges of credit reference based on big data technology in China. Mobile Networks and Applications, 27, 47–57. https://doi.org/10.1007/s11036-020-01708-y

[5] Xie, H. (2023). The applications of big data analysis in the credit business of

commercial banks. BCP Business & Management. https://doi.org/10.54691/bcpbm.v38i.4256

[6] Wang, F., Zhang, S., & Li, M. (2020). Influence of internet-based social big data on personal credit reporting. Asia-Pacific Journal of Convergent Research Interchange. https://doi.org/10.47116/apjcri.2020.07.05

[7] Wang, H. (2021). Credit risk management of consumer finance based on big data.

[8] Yang, X., Tao, Y., Li, G., & Jun, W. (2018). A theoretical credit reporting system based on big data concept: A case study of Humen Textile Garment Enterprises. In Proceedings of the 2018 International Conference on Big Data and Education. https://doi.org/10.1145/3206157.3206163

[9] Ransbotham, S. (2016). Using unstructured data to tidy up credit reporting. MIT Sloan Management Review, 57, 2.

[10] Gao, L., & Xiao, J. (2021). Big data credit report in credit risk management of consumer finance. Wireless Communications and Mobile Computing, 2021, 1–7. https://doi.org/10.1155/2021/4811086

[11] Cui, D. (2015). Financial credit risk warning based on big data analysis. [Unpublished manuscript].

[12] Sun, B. (2021). Bairong Zhixin: Big data and credit system construction. In Research Papers in Economics (pp. 357–400). https://doi.org/10.1007/978-981-16-1788-1_10

[13] Liu, X., Xu, Q., Wang, T., Ding, W., & Liu, Y. (2017). A credit scoring model based on alternative mobile data for financial inclusion. [Unpublished manuscript].

[14] Shi, Y. (2012). China's national personal credit scoring system: A real-life intelligent knowledge application. https://doi.org/10.1145/2339530.2339596

[15] Yiyuan, W., Yingfa, X., Yadi, L., Jiayue, Y., Xiaoping, Z., & Yuning, S. (2019). Big-data-driven model construction and empirical analysis of SMEs credit assessment in China. Procedia Computer Science, 147, 613–619. https://doi.org/10.1016/j.procs.2019.01.205

[16] Cheung, A., & Chen, Y. (2017). The transparent self under big data profiling: Privacy and Chinese legislation on the social credit system. LSN: Data Protection (Sub-Topic). https://doi.org/10.2139/SSRN.2992537

[17] Zhou, X., & Fu, Y. (2024). Does the big data credit platform reduce corporate credit resource mismatch? Evidence from China. Finance Research Letters. https://doi.org/10.1016/j.frl.2024.106133

[18] Cui, X., Sun, Y., Chang, X., Li, C., Zhang, G., Tu, D., Zeng, X., & Xiong, Y. (2016). A novel big-data-driven credit reporting framework for SMEs in China. In 2016 International Conference on Identification, Information and Knowledge in the Internet of Things (IIKI) (pp. 463–469).

https://doi.org/10.1109/IIKI.2016.46

[19] Mbah, G. (2024). Data privacy in the era of AI: Navigating regulatory landscapes for global businesses. International Journal of Science and Research Archive. https://doi.org/10.30574/ijsra.2024.13.2.2396

[20] Liu, C. (2018). Reflection on big data technology: Problems and countermeasures in "big data credit reporting" of internet finance in China. DEStech Transactions on Computer Science and Engineering.

https://doi.org/10.12783/dtcse/cmee2017/20058

[21] Sargeant, H. (2022). Algorithmic decision-making in financial services: Economic and normative outcomes in consumer credit. AI and Ethics, 1–17. https://doi.org/10.1007/s43681-022-00236-7

[22] Shittu, A. (2022). Advances in AI-driven credit risk models for financial services optimization. International Journal of Multidisciplinary Research and Growth Evaluation. https://doi.org/10.54660/.ijmrge.2022.3.1-660-676

[23] Segun-Falade, O., Ekeh, A., Odionu, C., & Chintoh, G. (2025). Challenges and conceptualizing AI-powered privacy risk assessments: Legal models for U.S. data protection compliance. International Journal of Frontline Research in Multidisciplinary Studies. https://doi.org/10.56355/ijfrms.2025.5.1.0036

[24] Wang, C., & Huang, R. (2023). FinTech-Bank partnership in China's credit market: Models, risks and regulatory responses. European Business Organization Law Review, 24, 721–755. https://doi.org/10.1007/s40804-023-00269-3

[25] Harvey, W. (2019). The impact of big data on credit scoring and alternative lending. EPH - International Journal of Science and Engineering.

https://doi.org/10.53555/ephijse.v5i3.267

[26] Schultz, J., & Crawford, K. (2013). Big data and due process: Toward a framework to redress predictive privacy harms. Boston College Law Review, 55, 93

[27] Feng, X., & Chen, L. (2022). Data privacy protection sharing strategy based on consortium blockchain and federated learning. In 2022 International Conference on Artificial Intelligence and Computer Information Technology (AICIT) (pp. 1–4). https://doi.org/10.1109/AICIT55386.2022.9930188

[28] Mangues-Bafalluy, J., Blanco, L., Javed, F., & Zeydan, E. (2024). Blockchain

and trustworthy reputation for federated learning: Opportunities and challenges. In 2024 IEEE International Mediterranean Conference on Communications and Networking (MeditCom) (pp. 578–584).

https://doi.org/10.1109/MeditCom61057.2024.10621302

[29] Zheng, F., Li, K., Tian, J., & Xiang, X. (2020). A vertical federated learning method for interpretable scorecard and its application in credit scoring. ArXiv, abs/2009.06218.

[30] Pingulkar, S., & Pawade, D. (2024). Federated learning architectures for credit risk assessment: A comparative analysis of vertical, horizontal, and transfer learning approaches. In 2024 IEEE International Conference on Blockchain and Distributed Systems Security (ICBDS) (pp. 1–7).

https://doi.org/10.1109/ICBDS61829.2024.10837430

[31] Song, Y., Ye, O., Ye, X., Yang, A., Liang, Y., & Liu, Z. (2021). A methodology of trusted data sharing across telecom and finance sector under China's data security policy. In 2021 IEEE International Conference on Big Data (Big Data) (pp. 5406–5412).

https://doi.org/10.1109/BigData52589.2021.9671857

[32] Müller, T., Zahn, M., & Matthes, F. (2023). Unlocking the potential of collaborative AI: On the socio-technical challenges of federated machine learning. ArXiv, abs/2304.13688. https://doi.org/10.48550/arXiv.2304.13688

[33] Zavolokina, L., Eckhardt, S., Sprenkamp, K., & Fernandez, J. (2024). Overcoming intergovernmental data sharing challenges with federated learning. Data & Policy, 6. https://doi.org/10.1017/dap.2024.19