WHIOCE PUBLISHING PTE. LTD.
PROVIDING
FIRST-CLASS SCIENTIFIC INFORMATION
FOR TOP SCHOLARS

2025 Volume 3, Issue 11
Educational Innovation Research
ISSN: 3029-1852 (Online)
ISSN: 3029-1844 (Print)

Research on Personal Data Security in the Context of Big Data

Genyuan Wang
Hainan Vocational University of Science and Technology, Haikou 571126, Hainan, China

Author to whom correspondence should be addressed.

Abstract: The deep integration of big data technology has accelerated the digital transformation of social production and daily life. As a core production factor in the digital economy, personal data's strategic value has become increasingly prominent, yet the accompanying security risks pose severe challenges to individual rights, social order, and even national security. Based on the circulation patterns of personal data and security governance practices in the big data environment, this paper systematically analyzes the core characteristics of personal data and potential security risks throughout its lifecycle. By examining authoritative data and typical cases from 2024 to 2025, it identifies prominent issues in current data security protection across technological, institutional, and governance dimensions. The study proposes a comprehensive protection framework encompassing technological innovation, institutional improvement, multi-stakeholder governance, special group protection, and localization of international experience. This framework provides practical pathways and theoretical references for achieving a balance between personal data security and data value development.

Keywords: Big data; Personal data security; Privacy protection

Online publication: Dcecmber 20, 2025

1. Introduction

With the rapid advancement of information technology and the internet industry, big data has become deeply embedded in various sectors including daily consumption, public services, and industrial upgrading. Personal data, as a core component, encompasses multidimensional content such as identity information, behavioral patterns, and health status [1]. The large-scale collection, storage, and analysis of such data have not only enabled precision services but also introduced significant security risks. According to regulatory statistics, 112 personal information leaks occurred in domestic government and enterprise institutions in 2024, involving 26.69 billion data entries. E-commerce and internet platforms were the most affected sectors. Data breaches and illegal trading not only infringe on individual rights but also risk triggering social trust crises, thereby constraining the development of the digital economy.

Against this backdrop, establishing an effective personal data security protection system has become a pressing challenge of our time [2]. Current data security protection faces multiple challenges, including inadequate technological adaptation, lagging institutional norms, and incomplete governance mechanisms. To address these issues, it is essential to adopt a holistic approach that spans the entire data lifecycle, integrating technical, legal, and managerial resources to create

a comprehensive protection framework. This approach ensures the safeguarding of individuals' legitimate rights while fully unlocking the value of data as a strategic asset.

2. Characteristics and security risks of personal data in the context of big data

2.1. Core features of personal data

Personal data in the big data era exhibits distinct characteristics that markedly differ from traditional data. The data is vast in scale and diverse in type, encompassing not only structured basic information such as names and ID numbers, but also unstructured and semi-structured data like web browsing records, audio-video content, and geographical location trajectories, sourced from multiple channels including internet platforms, mobile devices, and IoT sensors [3]. While data has low value density but high latent value, individual data points hold limited independent value. However, through big data mining and correlation analysis, core information such as user behavior patterns and consumption tendencies can be extracted, providing support for precision marketing and social governance. Data also demonstrates strong mobility, frequently flowing between different platforms, institutions, and regions. For instance, in e-commerce shopping scenarios, personal information must be shared among platforms, payment institutions, and logistics companies. This cross-entity flow significantly increases the complexity of data management and protection [4].

2.2. Security risks of personal data throughout its lifecycle

Every stage of personal data's lifecycle carries significant security risks [5]. In data collection, issues like excessive and covert collection remain prominent. Some institutions obtain consent through default selections and vague prompts, even resorting to illegal web scraping or data purchases, depriving users of control over their data. The 2025 report by the National Computer Virus Emergency Response Center reveals that while e-commerce apps have improved compliance since the implementation of the "Regulations on Network Data Security Management", "excessive collection of sensitive personal information" remains the top violation. Its proportion dropped from 27.11% in 2024 to 18.69% in 2025. Some platforms collect irrelevant data like biometric information and health status under the guise of "personalized recommendations", circumventing users' right to know through "one-click consent". In the first half of 2025, 32 non-compliant apps were fined between 1 million and 2 million yuan, with a leading fresh food e-commerce platform fined 1.8 million yuan for collecting allergy history data.

Data storage is vulnerable to technical vulnerabilities and management lapses, with issues like outdated database patches, weak passwords, and unencrypted storage making it susceptible to hacking. The lack of controls over data usage and sharing enables unauthorized misuse and illegal resale, fueling a black-and-gray market. Data analysis carries de-identification risks, as cross-referencing multi-source data can reveal personal identities. Furthermore, IoT firmware flaws, phishing Wi-Fi networks, and counterfeit apps exacerbate data breaches. With diminished user control over data and difficulty in tracing its flow, these factors collectively amplify security risks.

3. Prominent issues in current personal data security protection

3.1. Insufficient compatibility of the technical protection system

Current data protection technologies struggle to meet the complex demands of big data environments. Traditional encryption methods face challenges with massive data volumes, featuring low encryption/decryption efficiency, poor adaptability to unstructured data, and complex key management that could lead to data security breaches if keys are lost or stolen. Anonymization and desensitization techniques face security risks, as enhanced data analytics capabilities make traditional anonymized data vulnerable to re-identification. Access control systems contain vulnerabilities like improper permission settings and inadequate authentication mechanisms, increasing risks of internal overreach and external intrusions. Meanwhile, digital encryption technologies lag behind big data applications, further compromising

data transmission and sharing security. This gap is particularly acute in sensitive sectors like finance. In December 2025, Sichuan Rural Commercial Bank implemented an effective solution: collaborating with China Telecom Sichuan Company to establish the nation's first commercial OTN quantum encryption dedicated line. The system creates a triple-layer security barrier combining quantum key distribution, national cryptographic algorithms, and domestic chips, with SM4 algorithm as the encryption standard. After deployment, it supports 15 million daily encrypted transactions with encryption/decryption latency under 1 microsecond, 75% faster than traditional RSA algorithms. Within three months of operation, it successfully defended against 42 eavesdropping attempts and 17 database cracking attempts, scoring 97/100 (industry average: 81) in the People's Bank of China's 2025 evaluation. This technological integration not only validates the value of emerging solutions but also highlights the limitations of traditional systems.

3.2. Imperfections in institutional norms and legal safeguards

The legal framework for data security remains incomplete, with certain provisions being overly generalized and lacking detailed implementation guidelines. Ambiguous definitions, scopes, and classifications of personal data have led to inconsistent enforcement standards in practice [6]. Core legal principles struggle to be implemented effectively, as the informed consent principle has been distorted into standard contract clauses. Lengthy and obscure privacy agreements make data processing rules difficult for users to comprehend, while the widespread practice of "forced consent" undermines users' autonomy. Inadequate rights protection mechanisms result in challenges such as difficulty in proving violations, high litigation costs, and limited recourse channels when personal data rights are infringed, failing to create effective legal deterrence. Specialized protections for vulnerable groups like minors and the elderly, as well as specific sectors including finance, healthcare, and education, remain unaddressed, leaving data security boundaries unclear [7]. Notably, the "Regulations on Network Data Security Management" issued by the State Council in August 2024 (effective January 1, 2025) marked a significant breakthrough. It clarifies third-party sharing responsibilities for data processors, mandates reporting of data breaches within 24 hours, and raises the maximum penalty for "massive data leaks" to 2 million yuan, with direct offenders facing fines of 50,000 to 200,000 yuan. While these measures fill some institutional gaps, the overall system still requires further refinement.

3.3. Lack of coordination in the multi-stakeholder governance mechanism

Personal data security protection involves multiple stakeholders including governments, enterprises, and individuals, yet current governance mechanisms lack effective coordination. Government oversight suffers from fragmentation, with poor interdepartmental and cross-regional coordination. Insufficient regulatory efforts on cross-border data flows and platform data sharing have hindered collaborative governance. However, government efforts continue to intensify. The 2024 "Clean Net 2024" special operation investigated over 7,000 cases of personal information violations and arrested 12,000 suspects, targeting industry insiders who steal/sell data and black/grey market data. These actions effectively disrupted criminal networks stemming from data breaches, laying the groundwork for multi-stakeholder governance. Enterprises fail to fulfill their primary responsibilities, prioritizing commercial gains over data protection. Many lack robust data security management systems, professional talent, and technical capabilities, coupled with inadequate internal training that weakens employee awareness. Industry self-regulation remains underdeveloped, lacking unified standards and evaluation systems. Enterprises rarely share data protection experiences, hindering the formation of a self-disciplined ecosystem. Users' insufficient data protection awareness and skills — such as casually granting permissions for convenience, using weak passwords, clicking suspicious links, and connecting to insecure Wi-Fi, further increase data leakage risks.

3.4. Insufficient adaptation of special group protection to international practices

The protection of personal data for digitally vulnerable groups, including minors and the elderly, lacks targeted measures, and the localization adaptation of international best practices remains inadequate. Minors, due to their immature cognitive development, are prone to leaking personal information when using online services, and their data misuse may adversely

affect their physical and mental health. The elderly, with insufficient digital literacy, struggle to identify risks such as counterfeit apps and phishing links, making them a high-risk group for targeted fraud. In terms of international experience adoption, the EU's GDPR features "one-stop supervision" and "high fines" mechanisms that differ from China's market environment. The U.S. model, primarily based on industry self-regulation, cannot be directly replicated, and the regulatory framework for cross-border data flows lacks detailed operational procedures, which hinders the international adaptation of personal data security protection.

3.5. Imbalance between data value and security protection

Current data governance exhibits a one-sided tendency of either "overemphasizing protection at the expense of utilization" or "prioritizing utilization while neglecting protection", failing to achieve an organic balance between data value development and security safeguards. On one hand, excessive focus on data protection may restrict reasonable data flow and utilization, hindering innovative development in the digital economy. On the other hand, an overemphasis on data value at the expense of security protection could lead to infringement of personal rights and undermine the trust foundation for digital economic development. In scenarios such as public data openness and cross-border data flows, unclear data security boundaries and the absence of effective risk assessment and control mechanisms result in data being unable to fully realize its public value while lacking comprehensive security protection.

4. Optimization path of personal data security protection in the context of big data

4.1. Building a big data-adaptive technical protection system

Accelerate innovation and application of data protection technologies to enhance targeted and effective technical safeguards. Develop and promote emerging encryption technologies such as homomorphic encryption and quantum encryption, drawing inspiration from Sichuan Rural Commercial Bank's integrated model of "quantum key distribution + national cryptographic algorithms" to improve encryption/decryption efficiency for massive data volumes, achieving comprehensive protection for both structured and unstructured data. Simultaneously, refine key management mechanisms through multi-factor authentication and tiered key storage to ensure security. Optimize anonymization and desensitization techniques by combining differential privacy and data traceability technologies to strengthen data resistance against re-identification, reducing privacy leakage risks while maintaining data usability. Strengthen access control technologies by implementing granular permission policies based on data sensitivity and user roles, utilizing multi-factor authentication methods like biometrics and digital watermarking to prevent unauthorized access and data tampering. Promote blockchain and distributed storage technologies to build secure and trustworthy data storage and sharing platforms leveraging their tamper-proof and traceable characteristics, addressing leakage risks in centralized storage. Accelerate the widespread adoption of national cryptographic algorithms to reduce dependence on foreign cryptographic technologies and ensure autonomous control over core technologies. See **Table 1**.

Table 1. Personal data security technology protection system for big data

Technical module	Specific technical means	Direction of application	Safety goal
New encryption technology	Homomorphic/quantum encryption, national cryptography integration	data encryption storage	Enhance encryption and decryption efficiency, and protect all types of data
Anonymization and Desensitization Technology	Differential Privacy and Data Traceability	data sharing analysis	Enhanced anti-tampering recognition, balancing usability and privacy
Access control technology	Fine permissions, biometrics, digital watermarking	data access control	Prevent unauthorized access and data tampering

Shared storage technology	blockchain, distributed storage	data storage sharing	Resolve centralized leakage and enable traceable tamper-proofing
Basic technical support	National Cryptography Standard, Key Classification, Multi-factor Authentication	The underlying support of the entire system	Get rid of foreign technology dependence and ensure self-control

4.2. Enhancing data security systems and legal safeguards

Building upon the "Regulations on Network Data Security Management", we will strengthen the legal framework for data security by establishing a scientific institutional structure. This includes formulating detailed implementation rules for specialized data protection legislation, clarifying the legal attributes, rights attribution, and protection standards of personal data. Operational guidelines will be refined for data collection, storage, usage, and sharing, with clear criteria for identifying infringement and defining legal liabilities. A compensation mechanism that equally considers both emotional distress and economic losses will be improved. The implementation process of the informed consent principle will be optimized, with standardized privacy agreements presented in concise and visual formats to ensure users' rights to information and consent. A "true consent" standard will be established, prohibiting coercive or ambiguous consent practices. Rights protection mechanisms will be enhanced through simplified procedures, dedicated data rights protection agencies, and accessible complaint channels to reduce user costs. Law enforcement will be intensified to combat illegal activities such as data breaches and illicit transactions. Specialized security standards will be developed for sectors like finance, healthcare, and education, clearly defining boundaries for data collection and usage. Specialized protection systems will be established for digital vulnerable groups, including minors and the elderly.

4.3. Improving the multi-stakeholder collaborative governance mechanism

Establish a multi-stakeholder collaborative governance system integrating government oversight, corporate self-regulation, social supervision, and individual participation. Strengthen government regulation by consolidating the experience from the "Clean Web" special operation, clarifying departmental responsibilities, and establishing cross-departmental and cross-regional collaboration mechanisms. Utilize big data and artificial intelligence to build intelligent regulatory platforms for dynamic monitoring and risk early warning throughout data lifecycle, while formulating industry standards and evaluation systems to guide enterprises in standardized operations [8]. Reinforce corporate accountability by requiring enterprises to establish robust data security management systems, designate dedicated departments, enhance employee training, conduct regular risk assessments and audits, and improve emergency response plans to strengthen risk resilience. Promote industry self-discipline through association-led data protection conventions, corporate credit evaluation systems, and shared knowledge and technology to foster a self-regulatory ecosystem. Leverage social supervision by encouraging media and NGOs to expose data infringement cases, implementing whistleblower reward mechanisms, and conducting public education through online/offline training programs and public service advertisements to enhance data security awareness and protection skills, guiding users to adopt responsible internet practices and cautious authorization. See **Table 2**.

Table 2. Allocation of responsibilities among key actors in the multi-stakeholder collaborative governance framework for personal data security

Governance subject	Responsibility proportion	Core positioning
Regulatory level	35%	rule maker and regulatory leader
Self-regulation level	30%	Person responsible for accountability, core layer of execution
Level of social supervision	20%	Supervision of the Regulator and the Popularizer
Personal involvement level	15%	Rights defenders and end practitioners

4.4. Strengthening protection for vulnerable groups and localization of international practices

To address the alignment of international best practices with special populations, it is essential to achieve precise protection while promoting localized implementation of global experiences. Establish a tiered data protection system for minors, requiring platforms to adopt "guardian consent + child-friendly design" with default disabling of non-essential data collection permissions and restricted collection/sharing of sensitive information. Implement age-friendly privacy policies using voice prompts and simplified interfaces to inform seniors about data processing rules, strengthen supervision of elderly-targeted apps, and crack down on fraud schemes exploiting senior information. Drawing inspiration from the EU GDPR's "right to data portability" and "privacy impact assessment" mechanisms, optimize relevant clauses based on China's realities. Adopt U.S. industry self-regulation models to encourage voluntary standards from associations, establish a cross-border data flow monitoring system, refine outbound data security assessment procedures, implement cross-border data flow whitelisting and standard contract filing mechanisms, clarify corporate security responsibilities for cross-border data transfers, enhance international law enforcement cooperation, and develop cross-border data breach notification systems. Strengthen collaboration with major countries and regions to elevate the international compatibility of personal data protection standards.

4.5. Achieving dynamic balance between data protection and value development

Explore synergistic mechanisms between data security and value development to unlock data potential while ensuring protection. Establish a tiered data classification and protection system, categorizing personal data into general, sensitive, and core categories based on sensitivity levels. Implement differentiated protection strategies: enforce strict safeguards for core sensitive data while relaxing restrictions on general data to facilitate reasonable data circulation. Promote technologies like data sandboxes and federated learning to achieve "usable but invisible" data, enabling data sharing and value extraction while safeguarding privacy. Standardize public data openness and cross-border data flows by establishing review mechanisms and desensitization standards for public data, ensuring no personal privacy leaks. Improve security assessment and regulatory frameworks for cross-border data transfers, clarifying safety requirements and accountability. Develop fair data value distribution mechanisms to protect data subjects' legitimate rights during utilization. Through policy incentives and benefit-sharing approaches, encourage enterprises to innovate data applications in compliance, achieving a win-win scenario that balances individual rights protection, corporate innovation, and public welfare.

5. Conclusion

In the era of big data, safeguarding personal data security constitutes a systemic endeavor encompassing technological, institutional, and governance dimensions, which directly impacts individuals' legal rights, digital economy development, and social stability. Current challenges include inadequate technological adaptation, lagging regulatory frameworks, and insufficient collaborative governance. This necessitates concerted efforts from governments, enterprises, communities, and citizens to establish compatible technological systems, enhance legal safeguards, improve collaborative governance mechanisms, strengthen protections for vulnerable groups, localize international best practices, and balance data protection with utilization. Future strategies must keep pace with technological advancements and evolving scenarios, continuously optimizing protection measures to ensure the system evolves with the times. While unlocking the value of data as a key economic driver, it is crucial to fortify personal data security defenses and support the high-quality development of the digital economy.

Disclosure statement

The author declares no conflict of interest.

References

[1] Luo W, 2025, Research on Personal Data and Privacy Protection Measures in the Era of Big Data. Information and Computer, 37(1): 41–43.

[2] Cai J, Li Y, 2025, Challenges and Solutions in Student Personal Information Protection Under the Background of Smart Education. Teaching and Management, 2025(24): 57–64.

[3] Zhang Y, 2024, Legal Issues of Privacy Protection in the Big Data Era. Journal of Cultural Studies, 2024(9): 135–138.

[4] Xue W, 2024, The Operation Mode, Theoretical Dilemma and Protection Path of Personal Information in the Era of Big Data. China Maritime Law Research, 35(2): 103–112.

[5] Zhang Z, 2023, Optimization Path of Personal Data Rights Protection in the Context of Digital Economy. Industrial Innovation Research, 2023(16): 54–56.

[6] Zhang Y, 2022, Personal Electronic Information Security Protection in the Big Data Era: A Case Study of Smartphones. Science and Technology Information, 20(8): 19–21.

[7] Wang L, Liu H, 2021, A Preliminary Analysis on Personal Data Protection in the Era of Big Data. Information Network Security, 2021(S1): 90–93.

[8] Zhu Y, Li X, 2021, An Analytical Framework for Personal Data Privacy Protection in the Big Data Era. Journal of Information Science, 40(1): 165–170.