Proceedings of the 6th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/71/20241710

A preliminary discussion on data privacy and data security issues

Xian'an Chen

Hangzhou No.4 High School International School, Hangzhou, 310000, China

Moment61019@hotmail.com

Abstract. With the expeditious evolution of technologies such as the internet, big data, cloud computing, and artificial intelligence, data has ascended to the status of an invaluable and indispensable resource within the intricate tapestry of modern society. Yet, the labyrinthine ethical quandaries that intertwine with the processes of data collection, storage, processing, and application have progressively surfaced, capturing the rapt attention of academia, industry, and government as they grapple with this nascent challenge. However, with the extensive application of data technology, issues of data ethics have become increasingly prominent, posing unprecedented challenges to personal privacy, data security, social equity, and justice. This article aims to delve deeply into the connotation and extension of data privacy and data security, as well as its significant role in current society. By examining relevant research findings at this stage, analyzing the major challenges faced by data ethics and exploring corresponding strategies, this paper delves into the multifaceted challenges faced by data privacy and data security, highlights the inadvertent data collection by smart devices, the potential for re-identification of blurred and anonymized information through big data analysis, as well as the ongoing threats posed by hacker attacks, internal vulnerabilities, and the complexity of data circulation.

Keywords: Data ethics, data privacy, big data analysis, data security.

1. Introduction

In the information age, the vigorous development of new technologies such as big data mining and artificial intelligence has promoted technological innovation and economic development [1]. At the same time, the negative effects of big data applications have become increasingly prominent, with incidents and phenomena such as price discrimination based on big data, privacy data breaches, abuse of APP permissions, and the recommendation of harmful content emerging endlessly. The ethical values that humans have based on the traditional world of life are also facing significant challenges. The power of big data brings with it commensurate challenges. However, there is currently a lack of both institutional and legal constraints on the ethical use of big data, as well as a lack of a well-established ethical framework for big data usage to prevent and mitigate issues.

In recent years, data ethics has become a hot topic of concern for various sectors both domestically and internationally. As data has emerged as a strategic asset with significant social and economic benefits, it has also given rise to ethical issues. These include illegal data collection, dissemination, and misuse, as well as the unlawful acquisition and retention of personal data, data misuse, date monopolization, weakening of individuals' control over their data, unfair application, and biased data

DOI: 10.54254/2755-2721/71/20241710

guidance. This paper explores the reasons for current data privacy leaks and the practical problems faced by data security, and attempt to propose some solutions, which can strengthen people's awareness of protecting personal data privacy and promote modern technological innovation.

2. Current predicament faced by data privacy

2.1. Smart devices automatically collect data, often without user's knowledge

In the realm of smart devices, an intricate dance of data collection unfolds, often without the spectator's — the user's — explicit knowledge. These devices, ranging from smartphones to smart home appliances and intelligently connected automobiles, possess the capability to automatically gather user data, a process that frequently transpires beneath the user's cognizant threshold. For example, smart speakers may persist in eavesdropping and accumulating voice information, all while the user remains oblivious. Similarly, smart cameras might capture facial data without prior notification or consent. The data acquisition techniques employed by these smart devices are often characterized by a high degree of complexity, rendering it arduous for the common user to decipher their underlying mechanisms and data flows. This opacity facilitates a certain technical concealment of some acts of personal information collection that deviate from the norm, rendering them difficult to detect and trace by the users themselves. Besides, when collecting user data, some smart devices do not clearly inform users of the purposes, methods, scope, and frequency of information collection and use through privacy policies or other means. They also fail to provide users with clear options to allow or refuse. This results in users tacitly agreeing to data collection without fully understanding the situation. Moreover, the user often neglects privacy terms. Even when privacy policies are provided by devices, many users often choose to neglect them due to their lengthy and complex nature, or inadvertently select the "agree" option, thereby failing to fully realize that their data is being collected.

Once collected, automated data collection systems may use personal data for purposes beyond what individuals have consented to. This includes profiling, targeted advertising, and even influencing decision-making processes that directly affect individuals' lives. The lack of control over how and why their data is being used can erode individuals' trust in these systems, which can also enable extensive surveillance and monitoring of individuals' online and offline activities. This can infringe upon their right to privacy, especially when these activities are not related to any legitimate purpose or are not conducted with their knowledge and consent.

2.2. Big data analysis may re-mine blurred and anonymized information

Big data analysis may re-mine blurred and anonymized information, which mainly stems from the characteristics of big data technology, data mining capabilities, and potential connections between data.

Data aggregation and permanent storage: In the era of big data, various types of data are permanently stored and aggregated to form big data. This big data can be used repeatedly and permanently, providing a rich resource for data mining. Data mining technology can extract implicit, previously unknown, but potentially useful information and knowledge from large amounts of incomplete, noisy, blurred, and random data. Through operations such as intersection, reorganization, and association, big data mining can uncover hidden connections between data. Although privacy can be protected by blurring or anonymizing individual data, the correlation between data can emerge when various types of information are aggregated into big data, causing these originally unrelated pieces to become re-linked. Big data mining technology can process massive amounts of data through powerful computing capabilities and advanced algorithms, performing operations such as intersection, reorganization, and association on these information fragments to reveal their intrinsic connections. For example, by combining multi-dimensional data such as geographical location, shopping habits, and browsing history, analysts may be able to re-identify the real identity or interest preferences of anonymous users.

Deanonymization technology is an important aspect of big data analysis. This technology uses known information and algorithms to attempt to restore the original state of anonymized data. Research shows that even if data is anonymized during collection, in a big data environment, it is still possible to reveal

DOI: 10.54254/2755-2721/71/20241710

personal sensitive information through cross-comparison and comprehensive analysis with other data sources. This ability to deanonymize poses potential risks for big data analysis in re-mining blurred and anonymized information. The research of Professor Latany Sweeney from Harvard University is a classic case. He proved through experiments that only three pieces of information—age, city of residence, and gender—are needed to find a large number of matching personal records in publicly available databases. This discovery reveals the powerful ability of big data analysis in re-identifying anonymous individuals. Even if personal information is blurred or anonymized at the initial stage, in a big data environment, this information may still be re-mined.

In the business sector, big data analysis is widely used in user profiling. By analyzing multi-dimensional data such as users' shopping records, browsing behavior, and social media interactions, companies can build detailed profiles of users, including their interests, consumption habits, and social relationships. Even if this data is anonymized during collection, by analyzing the correlations and cross-comparisons between different data sources, companies can still identify users' real identities or characteristics. This ability makes big data analysis valuable in precise marketing and personalized recommendations, but it also raises concerns about user privacy protection.

3. The current challenges faced by data security

3.1. Hacker attacks and internal threats

Hackers invade systems through various means to steal personal sensitive information or damage and extort systems. Hacker attacks not only violate personal privacy but also directly threaten the integrity and reliability of data. Hacker attacks are criminal acts that are beyond the control of personal data processors, but personal data processors should make their utmost efforts to prevent the theft and misuse of personal information collected and exchanged for commercial purposes by their clients.

In the context of big data, hacker attacks pose another serious issue for data security. Hacker attacks typically refer to illegal intrusion into other people's computer systems or networks through technical means to steal, tamper with, destroy data, or achieve other illegal purposes. With the widespread application of big data technology, the means of hacker attacks have become increasingly complex and diversified, posing a significant threat to data security [2].

3.2. Internal vulnerabilities

Data security governance covers the entire lifecycle. For small and medium-sized enterprises, the cost and effort of data governance have significantly increased. Furthermore, businesses often focus more on data security during storage and usage, while giving less attention to data security across the lifecycle stages of collection, transmission, sharing, and destruction. These often become risk points in the data security governance process and are also considered the most significant vulnerabilities in the eyes of attackers [3].

Internal employees, driven by personal interests, may obtain sensitive customer information through methods like data export or manual photography and then sell it, making this a significant source of data leakage.

3.3. Complexity of data scale and circulation

The development of cloud computing technology has greatly improved the utilization of computing and storage resources, allowing enterprises to easily build large-scale data storage. However, the rapidly expanding data scale poses new challenges for data protection and governance.

In complex business scenarios and technical conditions, data is acquired and transformed at different locations, making tracking it highly uncontrollable.

In the context of big data, the data lifecycle has added stages such as sharing and trading, making data flow the "norm," while static data storage becomes the "exception." The implicit retention of information across multiple stages leads to difficulties in tracking and controlling data circulation [4].

DOI: 10.54254/2755-2721/71/20241710

Therefore, in a complex circulation environment, the primary requirement for data security is how to ensure the security of sensitive data such as important national data, corporate confidential data, and user personal data. With the implementation and advancement of initiatives such as information system integration and sharing, massive data resources are further shared and aggregated. Preventing data from being illegally copied, disseminated, altered, or otherwise leaked during usage and circulation has become another significant challenge [4].

4. Solutions and suggestions for current difficulties

4.1. Some suggestions for current dilemmas in data privacy

First, to strengthen safety awareness education for personnel.

Data security management is a long-term and complex task that requires the participation of all relevant personnel, especially professional technicians and managers in related fields, who are the core personnel in data security management and should enhance their professional and specialized skills. In terms of professional ethics, relevant personnel should have a sense of responsibility for data security management, actively participate in this work, and strictly follow requirements for standardized operations to avoid operational risks. In terms of professional skills, relevant staff should continuously learn knowledge and techniques related to data security management and effectively combine them with practical work to identify and resolve security issues [5].

Second, to strength technological innovation. Use encryption technology, blockchain technology, and other methods to ensure data security and traceability. Encrypt sensitive data to ensure its security during transmission and storage. Adopt strong encryption algorithms such as AES and RSA to prevent unauthorized access or theft of data. Implement strict access control strategies to ensure that only authorized users can access sensitive data. Implement methods such as Role-Based Access Control (RBAC) or Attribute-Based Access Control (ABAC) to restrict data access based on user roles and permissions. During data analysis and sharing, desensitize sensitive information to prevent personal information leakage. Techniques such as data masking and data camouflaging can be used to transform sensitive data into unrecognizable or irrelevant information. Establish a security audit and monitoring system to conduct real-time monitoring of data processing activities, record key operations and abnormal behaviors, and promptly discover and respond to data security incidents [6].

Third, to strengthening data transaction supervision since it is a crucial aspect of ensuring data security and compliance. To achieve this, we need to establish a robust data transaction supervision mechanism, clarify the regulatory bodies and their responsibilities, and formulate detailed rules and standards for data transactions. At the same time, we should enhance the supervision and management of data transaction platforms to ensure they have comprehensive security measures and compliance mechanisms in place, thereby preventing data leaks and misuse. Furthermore, we should intensify monitoring and enforcement efforts regarding data transaction behaviors, severely crack down on violations of laws and regulations, and create an effective legal deterrent.

To further enhance data transaction supervision effectively, the following specific suggestions can be adopted. Firstly, promote transparent operations of data transaction platforms and require them to disclose relevant information about the transaction data, including data sources, usage purposes, and transaction parties, to facilitate supervision by regulatory agencies and the public. Secondly, establish a whitelist system for data transactions, certifying and authorizing compliant data transaction entities to reduce the risk of illegal data transactions. At the same time, strengthen cross-departmental and cross-regional regulatory cooperation to form a joint force in combating illegal activities in data transactions. Lastly, improve the level of regulatory technology, leveraging advanced technical means such as big data and artificial intelligence to enhance monitoring and early warning capabilities for data transactions, promptly identify and address potential security risks. Through the implementation of these specific suggestions, we can further refine the data transaction supervision system, ensure the security and compliance of data transactions, and promote the sustainable development of the data economy.

DOI: 10.54254/2755-2721/71/20241710

4.2. Some suggestions for current dilemmas in data security

4.2.1. Enhance data anonymization and desensitization. During data collection, implement strict anonymization measures to remove or replace personally identifiable information, such as names, ID numbers, and phone numbers. Use techniques like hashing algorithms and generalization to ensure that the data cannot be directly linked to specific individuals. Desensitize sensitive information to reduce data sensitivity, making it impossible to easily restore to its original form while maintaining a certain level of usability [7]. Common desensitization methods include data perturbation and data noise addition.

4.2.2. Improve data encryption and protection technologies. Employ advanced encryption technologies to encrypt data during storage and transmission, ensuring data security at all stages. Implement both symmetric and asymmetric encryption algorithms to make decryption more challenging. Strengthen network security protection to prevent hacker attacks and data leaks by deploying security devices such as firewalls and intrusion detection systems to monitor and filter network traffic, detecting and addressing potential threats promptly [8].

4.2.3. Enhance data access control and permission management. Establish a strict access control mechanism to ensure that only authorized personnel can access sensitive data. Use technical means such as Access Control Lists (ACLs) and Role-Based Access Control (RBAC) to refine access permission management. Implement a robust monitoring and auditing system to track data access and usage in real-time, using log analysis and anomaly detection to quickly identify and mitigate data leaks and misuse.

4.2.4. Compliance with laws, regulations, and ethical standards. Strictly comply with relevant laws, regulations, and ethical standards during data collection, processing, and usage. Ensure data is sourced legally and used appropriately, avoiding infringement of personal privacy and rights. When collecting user data, clearly inform users of the collection purpose, usage methods, and protection measures, and obtain their explicit consent. Ensure that users have full knowledge and choice over their personal data. Regulations should clearly stipulate that data collection, storage, and usage must follow strict privacy principles, including data minimization, purpose limitation, transparency, and user consent. Furthermore, individuals should have the right to control their data, including viewing, correcting, and deleting it, as emphasized and protected tin data privacy laws [9,10].

4.2.5. Technological innovation and continuous improvement. Invest in research and development of privacy protection technologies, such as differential privacy and federated learning, to analyze data effectively while safeguarding user privacy. Regularly assess and update data privacy protection measures to promptly identify and address potential security vulnerabilities and risks. Continuously adjust and optimize privacy protection strategies and technical means as technology evolves and application scenarios change.

5. Conclusion

This comprehensive discourse meticulously examines the intricate web of data privacy and security challenges that have emerged amidst the relentless progression of technologies such as the internet, big data, cloud computing, and artificial intelligence. It underscores the dual-edged nature of these advancements, while propelling societal and economic growth, which have also given rise to myriad ethical conundrums that imperil personal privacy, social equity, and data security.

The paper delves into the covert nature of data harvesting by intelligent devices, often without users' informed consent, revealing a disturbing trend that undermines individual autonomy. Furthermore, it underscores the perils posed by big data analysis, which can meticulously re-identify obscured and anonymized information, posing grave threats to personal privacy. The ever-present specter of hacker attacks, coupled with internal vulnerabilities and the labyrinthine circulation of data, further complicate the landscape of data security.

DOI: 10.54254/2755-2721/71/20241710

In light of these multifaceted challenges, the discourse advocates for a holistic approach to data protection, integrating both technological advancements and non-technological measures. It emphasizes the imperative of bolstering data anonymization and desensitization techniques, employing robust encryption methodologies, and instituting stringent access control mechanisms. Moreover, it underscores the significance of adhering to legal frameworks and ethical standards, while fostering continuous technological innovation and improvement in the realm of privacy preservation.

In conclusion, this insightful discourse underscores the urgent need for collaborative efforts among policymakers, technologists, and stakeholders to navigate the delicate balance between technological progress and the safeguarding of personal data. It serves as a timely reminder that in the digital age, the preservation of individual liberties and privacy must remain paramount.

References

[1] Chen Yi. Enlightenment of EU's Big Data Ethical Governance Practice to China. Library and Information Work. 2020,64(03).

[2] Minxin Wu. Analysis of Data Security in the Background of Big Data, The 38th China (Tianjin) 2024 IT, Network, Information Technology, Electronics, Instrumentation Innovation Academic Conference, 2024-07-15.

[3] Wencong Han. Research and Analysis on Problems of Data Security Documents. Information Systems Engineering. 2024 (04).

[4] Shixi Wang, Zhang Liang, Li Jiaojiao. Data Security Protection in the Era of Big Data----Focusing on Data Security Governance. Information Security and Communication Security. 2020 (02)..

[5] Hongjie Zhang. Research and Practice of Data Security Technology for Mobile Sharing. Cyber Security. 2024,15 (02).

[6] Pengjuan Jin, Chuan Wang, Caiyi Huang. Ethics of artificial intelligence:the balance between privacy and data security. Shanxi Science and Technology Daily, 2024.4.25.

[7] Tao Yu. Research on Leak Detection and Desensitization Technology of Multi-source Privacy Data. Harbin Engineering University, 2023.4.1.

[8] Jie Dai, Biying Sun, Qiu Yu, Pengdong Lu. Application of Data Mining Technology in Network Intrusion Detection, Integrated Circuit Applications. 2024,41 (06) .

[9] Shuai Xu. Data Privacy Protection and Legal Responsibility:Challenges and Countermeasures under the New Situation. Legal Expo. 2024 (14).

[10] Jingyi Lu. Research on data privacy infringement litigants, Wuhan University, Issue 04, 2021.