Academic Education Publishing House

Philosophy and Social Science Vol. 3 No. 8, 2026

Ethical Privacy Protection System and Optimization Path under Big Data Mining

Yuxi Liu

College of St Hild & St Bede, Durham University, Durham, United Kingdom

Abstract: With the rapid evolution of big-data technologies, data mining has increasingly penetrated personal privacy, giving rise to markedly heightened risks of ethical anomie. By sorting out typical judicial cases and current institutional texts, this paper analyzes behavioral patterns and ethical roots of privacy infringements caused by data mining, and identifies structural defects including deficient informed consent, lagging technical protection, and fragmented supervision. Research indicates that constructing a three-tier collaborative system of "ethical norms-technical protection-institutional constraints" can effectively respond to the aforesaid risks. Optimization should be realized by embedding ethics throughout the whole process, advancing in-depth adaptation between technologies and scenarios, and establishing a co-governance pattern among governments, enterprises and users. The establishment of such a system helps achieve dynamic balance between exploitation of data value and protection of individual rights and interests, and provides operable governance references for the standardized development of the digital economy.

Keywords: Big Data Mining; Ethical Privacy; Data Protection; Privacy-Preserving Computation; Collaborative Governance

1. Introduction

In the era of the digital economy, big-data mining has been deeply integrated into core sectors including healthcare, finance, transportation and education, and has become a fundamental force driving social operation. Nevertheless, while creating tremendous economic value, data mining has triggered severe ethical-privacy crises. Typical cases released by the Supreme People's Court cover scenarios such as nearly 2.88 million patient records stolen via covert data interfaces embedded in medical registration systems, personal travel trajectories resold by insiders within industries, and precision fraud perpetrated using vaccine-appointment data. Approaches to infringing citizens' personal information are growing increasingly intelligent and chain-based, reflecting deep-seated defects within the current data-governance system. Existing privacy-protection frameworks show insufficient adaptability for data-mining scenarios, and tension between technical-supervision capacity and the speed of data exploitation keeps expanding. Systematically constructing a privacy-protection system that takes into account ethical norms, technical support and institutional constraints has become a pressing subject to be addressed in digital governance[1].

2. Ethical-Privacy Risks Brought by Big-Data Mining and Their Causes

2.1 How Big-Data Mining Erodes Personal Privacy

Big-data mining erodes personal privacy mainly through three paths. First, abuse of service interfaces: data service providers surreptitiously collect user information in the background by technical means beyond the scope of legitimate authorization. A typical example is that outsourced medical-software enterprises install data interfaces during system maintenance and continuously import personal information of registered patients into self-built databases, stealing a total of 2.878 million records. Second, abuse of position-based privileges: internal personnel leverage system-access permissions to query and sell citizens' information beyond authorized scope, giving rise to an "insider-driven underground economy". Third, circumvention by technical means, including forging ID photos to bypass real-name authentication systems and implanting Trojan horse programs to intrude vaccine-appointment platforms. Attackers hack official data-query

platforms and export large-volume sensitive data in batches[2].

These three erosion paths share common features: strong concealment and prominent scale effects. A single infringement often involves millions of pieces of personal information, and the acquired data can be secondarily exploited by downstream crimes such as precision fraud and doxxing-fueled online harassment, generating multiplied harm and seriously jeopardizing citizens' personal safety and property rights.

2.2 Internal Roots of Ethical Anomie in Data Mining

Internal roots of ethical anomie in data mining are multi-dimensional. First, reverse incentives within interest structures constitute the core driving force: the commercial value of personal information far outweighs costs for rule-breaking, generating persistent arbitrage impulses. In one typical case, industry insiders gained approximately 190 000 Yuan through illegal conduct. Compared with risks of criminal punishment, short-term returns proved highly tempting. Second, widespread weak awareness of data-ownership aggravates anomie risks. Both data controllers and users lack clear understanding of right boundaries for personal information, rendering authorization and consent a mere formality and giving rise to the prevalent practice of "one-click blanket consent". Third, hollowed-out organizational ethical norms also merit attention. Some enterprises fail to build effective data-access auditing mechanisms. Employees conduct long-term data operations exceeding authorization without surveillance, while organizations do not incorporate data ethics into compliance-culture development. Mutually reinforcing, these three roots form structural soil for ethical anomie in data mining.

2.3 Corresponding Author

Although top-tier legislations such as the Personal Information Protection Law and the Data Security Law have been enacted successively, current privacy-protection frameworks still exhibit obvious adaptive gaps for data-mining scenarios. Firstly, rule design lags behind technological evolution[3]. Existing regulations mainly target static links of "collection-storage-usage" of data, yet lack precise regulatory tools for privacy risks arising from associative inference and behavioral profiling in dynamic data-mining and analytical processes. Secondly, breaks appear within the whole-process regulatory chain. Data mining spans multiple subjects and links from data collection and analytical mining to derivative utilization. Current supervision tends to focus on initial collection, while lacking penetrating supervision over intermediate mining procedures and downstream application scenarios, forming a structural blind spot characterized by "heavy emphasis on both ends and neglect of intermediate links". Thirdly, cross-departmental coordination mechanisms are absent. Sector-specific data of healthcare, education, transportation, finance and other industries fall under different regulatory authorities. Divisions of responsibilities for ethical review amid cross-domain data flows remain ambiguous, enabling infringing acts to exploit regulatory loopholes.

3. Composition and Operation of the Ethical Privacy Protection System under Big-Data Mining

3.1 Ethical-Norm Layer Centered on Informed Consent and Data Minimization

The ethical-norm layer serves as the value foundation for the entire protection system. The principle of informed consent requires data controllers to clearly inform users of data purposes, mining scope and potential risks prior to initiating mining activities, and obtain active authorization from users rather than relying on default checkboxes within formatted clauses. For big-data-mining scenarios, informed consent should be further upgraded into a "dynamic consent" mechanism: users are entitled to withdraw authorization at any stage, and withdrawal shall not affect normal service access[4].

The data-minimization principle demands that only the minimum volume of data directly relevant to specific purposes be collected and processed in mining, and over-scope associative analysis be prohibited. Taking medical-data scenarios as an example, non-medical data such as patients' address books and location trajectories should be explicitly excluded from mining scope during construction of disease-prediction models. Moreover, the ethical-norm layer should incorporate the purpose-limitation principle: associative inference outputs derived from mining shall not

be adopted for commercial decisions beyond originally authorized purposes, so as to institutionally cut off transmission channels leading from data-mining outputs to privacy infringements.

3.2 Technical-Protection Layer Supported by Encryption, Desensitization and Privacy-Preserving Computation

The technical-protection layer delivers operable engineering implementations for ethical norms. Encryption and desensitization technologies conduct differential-privacy processing, K-anonymization and data desensitization upon raw data, ensuring that statistical analysis can be completed by mining models "without seeing original personal identifiers", and fundamentally mitigating harm severity once data leakage occurs[5].

Privacy-preserving computation represents core innovations within the technical-protection layer, covering technical paradigms such as federated learning, secure multi-party computation and trusted execution environments. Federated learning enables participating parties to conduct local model training and only upload gradient parameters instead of raw data, realizing "data remaining within domain while value circulates". Targeting typical infringement scenarios such as service outsourcers embedding covert data interfaces during system maintenance, the technical-protection layer should enforce mandatory deployment of data-access audit logs and least-privilege access-control mechanisms, implementing real-time early-warning and blockage for any over-authorized data-calling behavior. Synergy between the technical layer and ethical-norm layer translates subjective willingness for compliance into objectively verifiable technical constraints.

3.3 Institutional-Constraint Layer Tied to Multi-subject Collaborative Supervision

The institutional-constraint layer constitutes external safeguards for sustainable operation of the protection system. Faced with the complex ecosystem of multi-subject and cross-sector big-data mining, a single regulatory authority cannot achieve full coverage. A collaboratively linked regulatory network consisting of governmental regulators, platform enterprises, industrial associations and independent third-party audit institutions should be constructed.

At the governmental level, graded and classified regulatory regimes for data-mining activities should be established, and mandatory ethical review and filing procedures should be implemented for mining projects touching sensitive sectors including healthcare, education and finance. At the corporate level, independent data-ethics committees should be set up to conduct ex-ante assessment on value orientations of mining algorithms, and external institutions should be introduced to carry out regular data-security audits. At the industrial level, cross-sector ethical standard systems for data mining should be promoted to unify compliance benchmarks across diverse fields. Improvement of the institutional-constraint layer also relies on refinement of accountability-tracing mechanisms. Blockchain technology can be adopted to keep full-process records of data mining, facilitating precise identification and effective pursuit of subject liabilities so as to generate tangible deterrence against rule-breaking. See Figure 1.

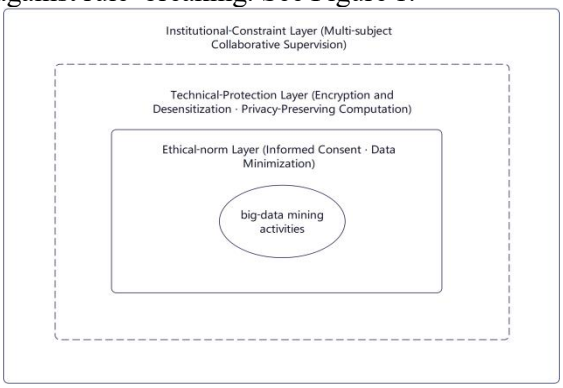


Figure 1 Framework of the Three-Tier Collaborative Privacy-Protection System under Big-Data Mining

4. Optimization Paths for the Ethical Privacy Protection System under Big-Data Mining

4.1 Strengthen Embedding of Ethical Norms throughout the Whole Process of Data Mining

The primary task for optimization paths is to shift ethical norms from ex-post remedies to ex-ante embedding across the entire workflow. At the data-collection stage, mandatory implementation of Privacy Impact Assessment (PIA) should be imposed. Data controllers are required to identify potential privacy risks and formulate preventive measures upon project initiation. At the mining-and-analysis stage, algorithm-ethical review mechanisms should be established to examine value orientations of

mining models that may generate discriminatory inferences or excessive profiling, closing structural loopholes of "rule-compliant data plus ethically improper mining".

At the output-application stage, usage scenarios for mining conclusions shall be explicitly restricted. Inferred data such as health-risk scores shall be prohibited from being applied to high-stakes personal decisions concerning employment and credit without granting parties the rights to know and appeal. Whole-process ethical embedding also calls for ethical-compliance incentive mechanisms. Ethical ratings of data-mining projects shall be incorporated into corporate credit systems to deliver positive guidance.

4.2 Promote In-Depth Adaptation between Privacy-Protection Technologies and Mining Scenarios

The core logic for technical optimization lies in breaking the myth of "technological neutrality" and advancing refined adaptation between privacy-protection capacities and concrete mining scenarios. For healthcare scenarios, federated learning facilitates cross-hospital training of disease-prediction models without centralized storage of original medical records, fundamentally eliminating leakage risks arising from data centralization. For educational-data scenarios, graded access control shall be implemented for high-value data such as student status and academic credentials, and differential-privacy technologies shall add statistical noise to query outputs to prevent precise identification through repeated queries.

For outsourced data-service scenarios — high-incidence zones for typical infringements — zero-trust architecture shall be rolled out, meaning that all internal access subjects are presumed untrusted. All data operations require multi-factor authentication and real-time comparison against behavioral baselines. Abnormal operations beyond baselines automatically trigger alerts and temporary privilege freezes. In addition, standardization and popularization of Privacy-Enhancing Technologies (PETs) shall be incorporated into industrial-access requirements, and technical-adaptation capacities shall be included in qualification reviews for data processing.

4.3 Build a Tri-partite Co-governance Pattern Linking Governments, Enterprises and Users

Constructing the co-governance pattern requires clarifying role positioning and interaction mechanisms for the three categories of subjects. Governments should transform from mere law-enforcers into institutional suppliers and incentive designers. Differentiated regulatory sandboxes for data mining shall be established to allow innovative enterprises to test novel mining technologies within controlled environments, while retaining rigid regulatory power over high-risk scenarios. For acts infringing citizens' personal information, the mechanism of incidental civil public-interest litigation shall be systematically popularized. As demonstrated by relevant judicial precedents, rule-breaking parties shall bear compound liabilities including paying compensation for public-interest damages, deleting illegal data and making public apologies.

Enterprises should internalize data-ethics governance into core competitiveness. A chief-data-ethics-officer system shall be introduced to advance simultaneous enhancement of internal accountability and external transparency for data-mining practices. For users, digital-literacy education and cultivation of privacy-right awareness shall transform passive data providers into right-holders who actively exercise entitlements. See Table 1 and Figure 2.

Table 1 Responsibilities System for Tri-Partite Co-Governance of Governments, Enterprises and Users

Subject	Core Responsibilities	Key Tools / Mechanisms	Reference Institutions
Governmental Regulators	Institutional supply, law-enforcement deterrence, cross-departmental coordination	Regulatory sandbox, graded and classified supervision, public-interest litigation	Incidental civil public-interest litigation (apology + compensation + data deletion)
Platform / Data Enterprises	Technical compliance, internal audit, ethical self-regulation	Data-ethics committee, zero-trust architecture, PIA assessment	Industrial data-security system construction
Data Subjects	Informed authorization, right	Dynamic consent withdrawal,	Articles 44-47 of the

(Users)	exercise, social supervision	complaint-reporting channels, digital-literacy education	Personal Information Protection Law of the People's Republic of China
Third-Party Audit Institutions	Independent evaluation, standard formulation, information disclosure	Ethical-review certification, public compliance reports	Ethical-standard systems formulated by industrial associations

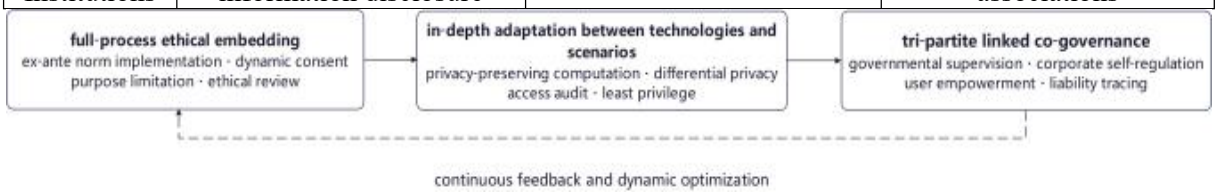


Figure 2 Closed-loop Mechanism for Optimization Paths of Big-Data-Mining-Related Privacy Protection

5.Conclusion

Tensions between big-data mining and privacy protection essentially reflect profound games between data valorization and protection of individual rights in the digital-civilization era. To resolve this dilemma, we should neither block data circulation and curb technological innovation in the name of privacy, nor allow unregulated privacy erosion in the name of development.

Construction and optimization of the three-tier collaborative system point toward a more systematic governance mindset: embed ethical constraints into the forefront of technical design, so that privacy protection becomes an endogenous attribute of data-mining architecture rather than an external compliance obligation. Judicial practices have demonstrated strong deterrent effects of combined application of public-interest litigation and concurrent punishment for multiple crimes. Nevertheless, ex-post sanctions remain passive lines of defense, and building proactive preventive whole-process governance mechanisms constitutes the fundamental solution. Only when technical rationality and ethical awareness are mutually embedded, and genuine consensus and joint efforts are forged among governments, enterprises and users, can we safeguard the bottom line of every citizen's data dignity amid in-depth advancement of the digital economy.

References:

[1] Press Office of the Supreme People's Court. The Supreme People's Court Releases Typical Cases on Courts Punishing Crimes of Infringing Citizens' Personal Information and Related Crimes [EB/OL]. (2026-05-08)[2026-09-22].https://www.court.gov.cn/zixun/xiangqing/499271.html.

[2] Zhang Kezhen, Wang Dong, Wang Yaling. Dilemmas and Solutions to Privacy Protection of Shared Data[J]. Finance & Economics, 2025, (12):49-60.

[3] Wang Fanyu, Sun Haoran. Real-World Problems and Practical Paths of Privacy Protection in the Big-Data Era[J]. Journal of China West Normal University (Philosophy & Social Sciences), 2025, (5):54-64.

[4] Liu Zhihong. Research on Privacy Protection and Data-Security Technologies for AI Large-Scale Models[J]. Software, 2024, 45(2):143-145+151.

[5] Mei Xiaying, Luo Ying. Legal Attribute of Data and Its Civil-Law Orientation[J]. Social Sciences in China, 2019, 40(1):82-99.