Personal Information Protection Dilemmas and Regulatory Adaptation in Generative Artificial Intelligence Applications

Yanghui Qu*
China University of Political Science and Law, Beijing 100088, China
*Corresponding email: 269140573@qq.com

Abstract

In the process of massive data collection and in-depth mining by generative artificial intelligence (AI). A large amount of personal information can be identified, posing numerous challenges in the field of personal information protection. Specifically, these challenges manifest as improper collection and abuse of personal information, increased risk of leakage, and expanded crises in group information security. Existing personal information protection rules struggle to adapt to the technical characteristics of generative AI, facing dilemmas such as the emasculation of the informed consent rule, difficulties in implementing the principle of data minimization, insufficient protection of personal information subjects' rights, and blocked remedies. To address these issues, this paper proposes path optimization solutions based on a risk-based governance framework: (1) strengthening the full-process application and supervision of Personal Information Protection Impact Assessments (PIPIA), and constructing a scenario-based hierarchical risk governance system; (2) optimizing existing rules in combination with technical realities, improving the hierarchical consent mechanism, and promoting the risk-oriented interpretation of the data minimization principle; giving play to the risk prevention role of procuratorial public interest litigation, and clarifying applicable standards and initiation conditions. The research aims to balance the technological innovation of generative AI with the protection of personal information rights and interests, provide theoretical support for the improvement of relevant rules and promote the healthy and orderly development of generative AI.

Keywords

Generative artificial intelligence, Personal information protection, Risk-based governance, Path optimization

Introduction

Artificial intelligence is profoundly transforming modes of production and life, impacting social progress significantly. Particularly, Generative artificial intelligence (GenAI), centered on deep learning, has brought great convenience. According to the *Report on the Application Development of Generative Artificial Intelligence (2025)* by the China Internet Network Information Center (CNNIC), China's GenAI application market landscape is gradually taking shape [1]. By June 2025, GenAI users in China reached 515 million, indicating broad development prospects [2]. However, as Nobel laureate in economics Joel Mokyr noted, throughout history, science and technology have driven economic development while also causing cultural anxiety [3]. The application of GenAI has similarly triggered unprecedented "intellectual anxiety". Due to its powerful information processing and content generation capabilities, GenAI relies inherently on mining and processing massive datasets [4]. For instance, ChatGPT's data scale has reached trillions of tokens, inevitably posing significant challenges to personal information protection during data collection and utilization. Unlike traditional information retrieval platforms, GenAI can automatically capture and process data in real-time, incorporating vast amounts of published or unpublished personal information into its corpus as training data, often without explicit knowledge or consent. This substantially increases the high risks to personal information security.

The extensive application of AI should serve human well-being and not become a persistent source of societal risk. Therefore, proactive thinking is required to regulate

this "Damocles' sword". On October 28th, 2025, the latest revision of the *Cybersecurity Law of the People's Republic of China* responded to AI governance demands for the first time, explicitly stating the need to "improve AI ethical norms and strengthen risk monitoring, assessment, and safety supervision". However, overall, specific norms and rules still require further refinement. Based on this context, this paper proposes targeted regulatory suggestions concerning the personal information security risks arising from GenAI applications, aiming to promote the healthy and orderly development of GenAI [5].

Definition and theoretical basis of core concepts

GenAI is a machine learning framework designed to generate diverse content by mimicking human creativity. It utilizes powerful computing capabilities to process and understand training datasets to produce content output. With the deepening integration of data elements, various types of personal information are presented in data form, and personal information is identified during data processing, bringing potential risks. In fact, models like ChatGPT and DeepSeek have faced investigations and bans by regulators in countries like Italy and Canada due to their personal information processing practices [6]. However, in pursuit of technological advancement, the industry often deliberately overlooks this issue, making personal information protection in GenAI applications an "elephant in the room". The *Report on Risk Perception and Information Transparency of Generative AI Users (2024)* released by the Nandu Digital Economy Governance Research Center shows that over 70% of surveyed users have some awareness of the potential risks of GenAI, particularly regarding personal information security, indicating clear concerns.

Risks in GenAI applications concerning personal information

Risks of improper collection and misuse of personal information

GenAI requires massive data as fuel. Whether engaging in deep communication with users or generating various graphic materials, data is the key factor enabling its high-level human-computer interaction. For example, Liang Wenfeng, founder of DeepSeek, pointed out that DeepSeek-R1 relies on large-scale reinforcement learning to stimulate model reasoning behavior, which demands enormous computational resources [7]. However, current public data resources are already insufficient, and the collection and use of public data alone would lead to model convergence and stagnation. Consequently, all GenAI service providers invariably target more personal information.

In practice, when large models acquire data, the scope and boundaries for collecting personal information are often completely vague and lack clear definition, leading to generalized risks of infringement on personal information rights. In terms of breadth, to pursue accuracy and richness in generated content, GenAI often collects information beyond what is necessary for providing basic services. Much user information not directly related to queries is incorporated into the corpus for analysis and learning [8]. In terms of depth, GenAI further re-identifies collected personal information through various complex intelligent analysis technologies, discovering correlations among vast amounts of personal data and mining its potential value, thereby aggravating the risk of rights infringement.

GenAI applications inevitably involve numerous human factors. Service providers, biased by commercial interests, may purposefully, consciously, and emphatically use personal information, facing risks of misuse. In practice, it is not uncommon for GenAI to acquire personal biometric information and infringe on personal information rights. For instance, in the case "Liao *v. A Technology and Culture Company - Dispute over Network Tort Liability*", the Beijing Internet Court ruled that the defendant company, using algorithms to fuse videos containing others' facial recognition features and expressions to achieve unauthorized "AI face-swapping", constituted infringement of the plaintiff's personal information rights and was liable for compensation. Furthermore, GenAI may form specific user profiles through personal online traces, consumption habits, and other information during use, further amplifying the risk of rights infringement through targeted responses.

Intensified and proliferating risks of personal information disclosure

The personal information collected and used by GenAI also faces serious leakage risks, with its security unguaranteed. Current large model technologies themselves contain vulnerabilities. Due to inherent flaws, personal information can be leaked due to improper

model design, storage system errors, etc. In March 2023, a technical vulnerability in ChatGPT's database led to the leakage of some users' names, email addresses, payment card information, and chat records. OpenAI immediately suspended service and issued a public apology. With technological advancements, in July 2025, ChatGPT was again exposed in a technological innovation experiment for leaking sensitive information, triggering global controversy [9]. As stated in Doubao's Privacy Policy: "Due to technical limitations and various potential malicious methods, even with our best efforts to enhance security measures, it is impossible to always guarantee 100% information security in the internet industry." Once a large model experiences security issues, personal information can be leaked on a massive scale like a flood, with unimaginable consequences. In July 2025, Amazon Q (a GenAI assistant developed by Amazon) was found to contain malicious code modification instructions capable of destroying systems and stealing large amounts of personal information. These high-risk instructions successfully passed Amazon's review and were downloaded hundreds of thousands of times. Fortunately, after exposure, Amazon promptly removed the plugin, avoiding serious consequences. However, this incident still aroused deep public concern about personal information disclosure in large model databases. Former Google CEO Eric Schmidt once pointed out the serious proliferation risks in AI.

Although major GenAI service providers guarantee protection of collected personal information through methods like encryption, fundamentally solving the problem is difficult, and the risk of disclosure remains objectively present. For example, a researcher once asked ChatGPT to repeat a word like "poem". After outputting a certain number of repetitions, the model suddenly began outputting nonsensical text containing large segments suspected to be from its training data, including someone's email signature and contact information. Mira Murati, former CTO of OpenAI, stated that ChatGPT must be used cautiously, and sensitive personal information should not be uploaded to avoid leakage and losses. Once personal information is leaked, it is highly likely to spread uncontrollably in cyberspace, leaving individuals virtually "exposed". The existence of GenAI exacerbates this harm, and infringement on personal information rights may continuously expand. As mentioned, even non-sensitive personal information can be reaggregated by GenAI's powerful computing power for information reconciliation.

After connection, it can reveal certain disclosures and then infer more comprehensive personal information. Some scholars therefore propose "derivative privacy data", where low-sensitivity personal information can also become identifiable to individuals with the help of algorithms. Through GenAI's deep analysis of personal information, its generated content becomes highly personalized, especially in user-model dialogues, directly exposing personal preferences and behavioral habits. Based on the "long-tail effect", after such personalized information is mined and utilized, wrongdoers can analyze specific individual profiles and carry out targeted illegal activities, including criminal acts like stalking, harassment, identity theft, and telecom fraud.

Group crisis in personal information protection

Literally, personal information seems to concern specific individuals. Article 4 of China's Personal Information Protection Law (PIPL) also reflects the legislative intent to protect individuals from easy identification in the online world. Individual damages in personal information infringement are often slight, leading to less public attention. However, with GenAI development, the value of personal information has transcended the individual category, and its associated risks affect the public interest of society [10]. Combining the perspective of Dutch scholar Anuj Puri, we can recognize that in the data processing landscape under the GenAI wave, the importance of individuals is gradually diminishing, with the core focus shifting to groups.

On one hand, when GenAI collects and uses personal information, the volume involved often reaches tens or hundreds of millions, and affected users can be broad groups. For example, in June 2023, OpenAI was accused of unauthorized use of up to 300 billion words of personal information for training, covering medical records, children's information, and many other sensitive data types [11]. Preliminary estimates suggest millions of affected users, potentially up to 3 billion. This shows that the "public" nature of personal information is infinitely magnified due to the large-scale frequency of processing and risk spillover effects of GenAI. Once misused, it not only damages individual rights but can also lead to mass incidents, systematically impacting the overall societal information ecology.

On the other hand, with GenAI's powerful data analysis capabilities, massive personal information is automatically extracted by large models after aggregation and desensitization. Some scholars note that the value of personal information shifts from "individual profiling" to "group patterns". If improperly used, this could lead to risks like group discrimination and social crises with the widespread application of GenAI. Additionally, in this context, even users who confirm they have not provided extensive personal information may be included in groups and subjected to privacy speculation. American scholar James Rule once vividly stated: "No matter how personal the information, we often depend on others to benefit from or suffer from the general gains and losses of privacy. If nearly everyone around me believes their conversations are monitored and alters their behavior, then even if I am sure my conversations are safe, something vital has been lost in public life."

Luciano Floridi, founder of information philosophy, once said: "In the big data era, most people are sardines. A sardine might think the fishing net is for itself, but it is not - the net's target is the entire school. Therefore, to save sardines, you must protect the school." Personal information possesses dual attributes of individual and public value. However, in GenAI application scenarios, the public attribute is magnified, and personal information is, in a sense, dissolved into the group concept. Long ago, scholars proposed "group privacy theory" and advocated "group data empowerment", but current Chinese law still focuses on protecting individual information rights. Mechanisms for protecting group information rights and interests remain underdeveloped, leaving the challenge posed by GenAI looming over social groups.

Dilemmas of existing personal information protection rules in the GenAI context

Samuel Warren and Louis Brandeis in the United States pointed out the obvious tension between individuals' desire to enjoy new technology benefits and their instinct to avoid privacy intrusion. In the digital age, balancing personal information use and protection remains the core issue. Protecting personal information is the starting point of China's data protection legislation. However, the emergence of GenAI challenges the existing normative framework, necessitating timely improvement. The *Interim Measures for the Management of Generative Artificial Intelligence Services*, effective August 2023, regards respect for personal information rights as an important principle but largely reiterates relevant requirements from PIPL. The national standard Cybersecurity Technology - Safety Specification for Data Labeling of Generative Artificial Intelligence (GB/T 45674-2025), effective November 1st, 2025, specifies data labeling requirements for both prompt and response information. However, it remains essentially a preliminary example from the service providers' perspective and lacks further discussion on personal information protection across different scenarios. Therefore, the existing legal system is incomplete in regulating personal information application in GenAI, and relevant departmental regulations have not proposed more practical, targeted governance measures, leaving this amplified risk even more uncontrolled.

The formalistic dilemma of the informed consent rule

The informed consent rule is the core criterion for personal information processing. The *Civil Code* stipulates that individual consent is a circumstance allowing personal information processors to avoid civil liability. PIPL regards "obtaining individual consent" as the primary legal basis for processing personal information. However, in the big data era, this rule often becomes a mere formality, especially with GenAI's use of personal information highlighting its formalistic dilemma [12].

On one hand, "informing" is not truly executed. In practice, GenAI typically provides notice during initial user registration via user agreements, particularly privacy policies. No further prompt pop-ups appear during subsequent actual use, fulfilling the obligation through a one-time, formalistic notice. Such a "privacy policy" obviously cannot exhaust all details of personal information collection and use. In contrast, Article 17 of PIPL imposes significantly higher requirements for the duty of notification. Furthermore, as mentioned, GenAI can extensively collect data via web crawlers to build corpora [13]. The personal information obtained in this process may come from various platforms, even undergoing multiple rounds of transformation, making it difficult for the large model itself to accurately trace and verify sources. Informing each individual data subject is not practically feasible. The "neural network" mechanism of GenAI data processing is extremely

complex, often not fully explainable even by designers and developers, objectively lacking conditions for adequately fulfilling the notification obligation. Additionally, the training process of GenAI datasets is highly automated and cannot interact timely with data subjects. Having large staff manually participate in notification would incur massive costs, contradict commercial efficiency and potentially hinder technological development. For data subjects, frequent notifications from processors would become burdensome and inconvenient over time.

On the other hand, user "consent" is not effectively obtained. GenAI privacy policies are often lengthy and complicated. For instance, DeepSeek's policy exceeds 10,000 characters, ERNIE Bot's exceeds 12,000, and Doubao's exceeds 13,000, with large portions bolden or highlighted. Users typically skim or ignore them. Even if read carefully, boundedly rational individuals cannot fully comprehend key points or adequately assess the risks and consequences of processing activities. Under such circumstances, users often have no choice but to "agree", otherwise they cannot access the service. This does not constitute genuine consent but rather passive acceptance. Moreover, GenAI dataset training is often continuous, with personal information constantly and repeatedly analyzed and used. The lack of transparency and openness in the training of corpus models prevents users from fully understanding key details, such as how their information is processed. As a result, they are unable to assess whether their data is being used and protected in a reasonable manner, which considerably weakens the validity of any initial "consent" given. More specifically, although Article 13 of PIPL stipulates a series of statutory exceptions where processors may process information without consent (e.g., "statutory duties", "public interest"), most are inapplicable to GenAI application scenarios. Although Section 2 of PIPL specifically regulates sensitive personal information processing, emphasizing the special requirement of "separate consent", in practice, such separate consent is never specifically obtained during applications. Even if obtained information is not initially sensitive, as mentioned, GenAI's powerful inference ability can leverage inter-information correlations to transform originally non-sensitive information into sensitive information. Whether "separate consent" is then needed or obtainable remains contentious [14].

The implementation dilemma of the data minimization principle

The data minimization principle requires that the collection and processing of personal information be limited to the minimum scope necessary for achieving specific, explicit, and legitimate purposes. However, the technological characteristics and application demand of GenAI render this principle difficult to implement in practice.

First, the "necessity" boundary is blurred. GenAI's performance heavily relies on data volume and diversity. To enhance output quality, richness, and accuracy, developers have inherent incentives to collect as much data as possible, including various types of personal information. This conflicts with the minimization requirement. Defining the "minimum scope necessary" for the broad goal of "improving service quality and user experience" is challenging, providing room for data over-collection.

Second, the "purpose limitation" principle is weakened. Once personal information enters GenAI's training corpus, it may be used for various unforeseen generation tasks, deviating from the initial collection purpose. The repurpose and secondary use of data are difficult to monitor and constrain.

Third, data retention and deletion face technical barriers. The principle also requires deleting or anonymizing personal information once the purpose is achieved or when retention is no longer necessary. However, large language models learn by adjusting billions of parameters; information is diffusely stored within the model rather than as discrete, deletable entries. As researcher Timnit Gebru once noted, "AI systems cannot forget like humans; old information remains stored in model parameters, albeit with reduced priority. To eliminate memory, the model must be retrained". For cost reasons, retraining a large model equates to rebuilding it, with almost no practical feasibility, leaving the right to erasure shelved.

Insufficient protection for data subject rights

Under the information society's legal order, there is a serious status imbalance between individuals and personal information processors. PIPL aims to enhance protection of personal personality interests and individuals' ability to control their information, eliminating information asymmetry through rights like erasure, correction, access, and copy. However, GenAI's

demand for personal information far exceeds traditional platforms like search engines. After collecting vast amounts of personal information from different subjects, various data points attach and mix - what some scholars call a “data soup”. Consequently, users struggle to trace the flow of their original information, hindering effective exercise of their rights and challenging the core purpose of safeguarding personal information rights. Moreover, even if users claim their rights, the cost for GenAI developers to fulfill obligations is prohibitively high, leading to formalistic responses with little substantive effect.

Blocked pathways for relief of personal information infringement

When GenAI unreasonably uses personal information, infringing upon rights, it is practically difficult for users to obtain protection through legal means. According to an announcement by the Cyberspace Administration of China (CAC), as of November 2025, 611 GenAI services have been filed in China, most belonging to specialized “little giant” enterprises, while well-known domestic services like Tencent Yuanbao, ERNIE Bot, and Tongyi Qianwen are products of large, listed companies such as Tencent, Baidu, and Alibaba. In contrast, competition for data elements among related enterprises is intensifying.

The litigation path for a weak individual is extremely difficult, with little chance of success alone. Due to the prevalent “algorithmic black box” in GenAI rules, internal mechanisms are unexplainable and incomprehensible. The complexity of personal information processing activities leaves data subjects completely “in the dark”, unable to confirm a legal causal relationship between action and result, increasing the concealment of infringement. After realizing rights infringement, the investigation and evidence collection phase faces severe difficulties. GenAI typically does not provide users with transparent access methods and deliberately hides underlying decision-making logic. Although PIPL grants natural persons rights to access and copy information for evidence collection, defendants often cite high costs to refuse in judicial practice. Article 8.1 of the Information Security Technology - Personal Information Security Specification (GB/T 35273-2020) and other relevant provisions provide some legal basis for this, further highlighting evidentiary challenges. Claiming rights entails significant time and financial costs with difficulty obtaining comprehensive favorable materials. Even with reversed burden of proof, some scholars note that the lack of algorithmic interpretability makes traditional tort principles difficult to apply effectively in intelligent algorithm scenarios.

Optimization pathways based on risk governance

To better address challenges to personal information rights protection in this evolving era, many scholars propose shifting from traditional rights protection paths to risk control paths, further balancing protection and utilization. The risk control path emphasizes identifying, assessing, distributing, and managing personal information protection-related risks, taking effective preventive measures to minimize impact. On one hand, this requires personal information processors to undertake more critical risk management obligations rather than shifting risks to individuals, effectively strengthening practical needs for personal information security protection. On the other hand, it pursues optimal risk combinations, focuses on probabilities of potential adverse consequences, abandons absolute pursuit of zero-risk, and promotes rational technological progress. This protection path is reflected in laws, regulations, and protective rules of many countries or regions, such as the EU’s General Data Protection Regulation (GDPR). China’s PIPL also responds to this, fully reflecting a risk-based governance framework in its overall protection mechanism.

However, GenAI development brings rapid changes, revealing shortcomings in this governance path. Risk means “possible danger”, but what kind of risk requires addressing? Some scholars note the difficulty in predicting whether, when, how, and whom a boulder that may fall from a cliff will affect. Professor Zhang Lingling points out that current criteria for judging risks remain vague, classification determinations are inconsistent, and evaluation mechanisms often become formalistic, failing to effectively cover all dangers personal information faces in GenAI application scenarios [15]. In short, GenAI’s technical complexity imposes higher requirements on risk management, necessitating further improvement, optimization, and flexible adjustment of governance strategies to meet comprehensive challenges posed by GenAI to personal information protection.

Strengthening the application of personal information protection impact assessment in GenAI

PIPIA is considered the most direct and concrete application of risk management methods in personal information protection. The Information Security Technology - Guide for Personal Information Security Impact Assessment (GB/T 39335-2020) clarifies that this system needs to test the legal compliance of personal information processing activities, judge risks to data subjects' legitimate rights and interests, and evaluate the effectiveness of protective measures. Similarly, in ecological environmental protection, the Environmental Impact Assessment (EIA) system focuses on investigating, predicting, and evaluating potential adverse impacts in advance, becoming an inviolable red line for construction projects over years of development. In contrast, PIPIA still requires further improvement. For example, during the "Clean Net 2025" special operation by public security cybersecurity departments, an AI service technology company faced administrative penalties for processing sensitive personal information without conducting a PIPIA. This "first PIPIA case" sounded another security alarm. Article 55 of PIPL lists a series of preconditions for applying PIPIA in China. GenAI service development inevitably involves automated decision-making, aligning with these application conditions.

(1) Building a scenario-based risk classification and hierarchical governance scheme

Risks for personal information in GenAI applications vary across different usage scenarios (e.g., general Q&A, content generation, personalized recommendation, professional services). A one-size-fits-all governance approach is inadequate. Therefore, based on risk levels, personal information types (general vs. sensitive), and application scenarios, a scenario-based risk classification and hierarchical governance scheme should be established. For high-risk scenarios (e.g., processing sensitive personal information like biometrics, health data, or involving automated decision-making with significant impacts), stricter governance measures are required, such as mandatory PIPIA, higher transparency requirements, and enhanced rights protection mechanisms. For medium and low-risk scenarios, relatively relaxed compliance requirements may apply to avoid hindering technological innovation while ensuring basic security. The recently implemented national standard GB/T 45674-2025 represents a preliminary attempt at scenario-based governance but needs further refinement regarding specific classification standards and differential measures.

(2) Strengthening public participation and external supervision

Risk perception varies among individuals, especially concerning GenAI involving user personal information rights, necessitating introducing external supervision into impact assessment analysis. Public participation is important in China's ecological environment protection legislation, and the PIPIA system can draw lessons. After all, the public, as stakeholders, can restrict GenAI enterprise behavior, help identify undiscovered security risks, and contribute to more complete and objective impact assessment reports. However, trade secrets of GenAI enterprises are crucial. Excessive public participation may undermine corporate confidentiality mechanisms and even increase personal information disclosure risks. Therefore, appropriate stipulations should balance enterprise realities.

Article 58 of PIPL provides for establishing a personal information protection supervision and management committee composed of external members to "supervise personal information protection", but it has long been shelved without genuine development. On September 12, 2025, the CAC issued the *Provisions on the Establishment of Personal Information Protection Supervision Committee for Large-Scale Online Platforms (Draft for Comment)*, whose Article 3 indicates a move toward concretizing this mechanism. In the future, this committee's role in PIPIA oversight should be strengthened, conducting independent reviews of assessment reports for high-risk GenAI services, raising opinions, and monitoring rectification. This combines external professional oversight with internal corporate responsibility.

(3) Improving the whole-process dynamic mechanism

Although Article 55 of PIPL indicates PIPIA is conducted "beforehand", simply limiting it to prior review is insufficient for the long-term governance of GenAI to ensure risks remain within acceptable ranges. Therefore, dynamic and continuous review of personal information risks should be conducted, including pre-assessment, in-process review, and post-tracking. First, for pre-assessment, GenAI service providers should review whether planned personal information processing

activities comply with laws and regulations, identify risks, and take corresponding security measures. Second, for in-process re-examination, due to GenAI's autonomy and unpredictable algorithmic black-box nature, risks may not remain static during operation. Timely re-examination, when necessary, continuous monitoring throughout the GenAI lifecycle, judging whether safeguards adapt to risk changes, and regular evaluation of impacts on personal information rights are required. Third, for post-incident follow-up, enterprises should be required to file PIPIA reports with competent authorities and regularly review PIPIA effectiveness, continuously improving protection. The PIPIA system and the compliance audit obligation under PIPL Article 54 should not be isolated; both compliance assessment and risk assessment deserve equal attention, enhancing GenAI service providers' awareness and technical capabilities for personal information protection.

Optimize interpretation of protection rules considering GenAI development

(1) Optimizing the informed consent rule within the existing framework

The informed consent rule's application is challenging under the GenAI wave and should be improved and optimized within the existing framework to avoid long-term ineffectiveness. Strict adherence may impose burdens on both service providers and users. To better promote technology and protection of rights, a tiered consent rule can be constructed based on a risk-based understanding of notification-consent. PIPL already embodies this by roughly dividing personal information sensitivity. Many scholars propose classification schemes, e.g., Professor Zhu Xiaofeng's suggestions on sensitive private information, sensitive non-private information, and general personal information [16]. By presenting consent rule systems with varying protection intensities, personal information utilization thresholds can be differentiated. On one hand, this allows setting notification obligations of varying intensity for GenAI, reducing costs. On the other hand, it reflects focus on vital personal interests. However, with rapid technological development, information sensitivity can change quickly, while existing classification is more like fixed "labeling". GenAI's powerful inference ability may cause originally low-sensitivity information to yield unexpected processing results in different scenarios. Therefore, tiered consent strategies should be flexibly formulated, reserving governance space for GenAI's future evolution. Additionally, GenAI service providers' obligation to create transparent environments should be considered comprehensively. For legally collected general personal information, publicity can inform the public of collection sources and processing rules, acknowledging users presumed consent based on investigated service scenarios. For high-sensitivity information, service providers should clearly describe it on separate notification interfaces, optimize notification forms, and provide users with more convenient response mechanisms.

(2) Promoting risk-oriented interpretation of the minimization principle

The minimization principle regulates personal information processing through relevance, minimization, and proportionality sub-principles. However, GenAI's operating mechanism limits this principle in practice, necessitating reinterpretation emphasizing relevance in collection, minimization in mining, and proportionality in risk to promote its application centered on risk control. First, when GenAI collects personal information, it must still be limited to service purposes, meeting relevance requirements, avoiding over-collection. During pre-training, unnecessary personal information types should be eliminated timely, with relevant departments supervising and ordering rectification/punishment for non-compliant collection lists. Second, GenAI's powerful deep processing ability should be limited to ensure reasonable understanding depth vertically, generating content around user demands, avoiding over-mining, and meeting information processing minimization. This can be achieved by judging required processing depth for different GenAI models in different scenarios through repeated simulation experiments, processing information based on understanding user intentions, limiting GenAI's mining as much as possible with existing technology to avoid potential risks, and introducing adversarial samples for training, forcing models to learn and generate more essential content. Third, GenAI's collection and processing should meet proportionality requirements, matching service quality and generated content accuracy, and aligning with risks of infringement to personal information rights. Processing activities should integrate with PIPIA reports, carefully considering factors like interest balance and

risk assessment to achieve dynamic balance between rational utilization and effective protection.

(3) Strengthening protection for exercising data subject rights

Rights like erasure, access, and copy are also contentious under GenAI applications - technically difficult and costly. However, service providers cannot arbitrarily violate legal provisions and ignore user requests. Regarding the right to erasure, GenAI should, through technical upgrades, block relearning of preliminarily deleted personal information, deploy effective filtering mechanisms to shield reappearance of relevant information during output, achieving "prohibition of generation" even if unable to "forget", preventing direct reproduction of deleted information in generated content. Simultaneously, user application channels should be optimized, providing alternative methods beyond email contact, facilitating users in proactively avoiding rights infringement risks. Regarding access and copy rights, if involving large amounts of complex, deep-seated information, users need to provide clear reasons to the service provider, who can then decide by comprehensively considering various factors with reference to GB/T 35273-2020 provisions. Subsequently, relevant departments should list circumstances where GenAI service providers may not provide copies, such as based on risk level, technical feasibility, and cost, preventing arbitrary rejection of user access requests on vague grounds and protecting users' right to know.

Emphasizing the risk-preventive role of procuratorial public interest litigation in personal information protection

As mentioned, when personal information rights are seriously infringed, affecting public interests, individuals may struggle effectively using legal remedies against GenAI. Therefore, the role of procuratorial public interest litigation should be emphasized to improve governance efficiency in personal information protection. Procuratorial organs' advantages in discovering clues and mastering evidence help better address public interest protection challenges under the GenAI wave. PIPL Article 70 also establishes a personal information protection public interest litigation system, providing important legal support for safeguarding public interests. However, judicial application requires further refinement to effectively translate institutional advantages into judicial effectiveness for protecting personal information rights.

On one hand, PIPL states that when a processor infringes upon "the rights and interests of a large number of individuals", procuratorates may file lawsuits. Once GenAI applications infringe on personal information rights, they often involve numerous users, meeting the subjective quantitative standard of "large number". However, lacking clear legal definitions, grassroots procuratorates may hesitate to prosecute, leading to further expansion of infringement. Therefore, the Supreme People's Court should issue relevant judicial interpretations, or the Supreme People's Procuratorate should issue regulations clarifying the "large number" threshold. The Supreme People's Procuratorate's *Notice on Implementing the Personal Information Protection Law and Promoting Public Interest Litigation for Personal Information Protection* points out that procuratorial organs at all levels should focus on protecting personal information of over one million people. This provision guides prosecution of personal information protection public interest litigation and should inform future definitions of "large number", with quantitative standards appropriately lowered, especially in GenAI-related personal information infringement fields, to recognize and protect group information and corresponding group rights, fully protecting personal information rights.

On the other hand, PIPL's wording "infringes" obviously implies processing activities have caused actual damage. However, China currently lacks cases where GenAI causes large-scale user personal information infringement; more exist only as potential urgent risks, lacking legal basis for initiating procuratorial public interest litigation. Due to GenAI's astonishing propagation speed, untimely response could cause irreversible damage. Therefore, when GenAI disregards PIPL principles and constantly probes the edge of infringement, with risks exceeding estimated thresholds, procuratorial organs should also be allowed to intervene via public interest litigation to perform duties. Article 4 of the *Public Prosecutorial Public Interest Litigation Law of the People's Republic of China (Draft)*, which began soliciting public opinions on October 28th, 2025, stipulates that procuratorates should "prioritize protection and emphasize prevention" in handling public interest litigation cases, which can serve as a basis for

preventive public interest litigation against GenAI personal information infringement risks. However, this principled provision is insufficient; quantitative standards for risk damage should be explored, and both PIPL and the Public Prosecutorial Litigation Law should clarify that procuratorial organs may file lawsuits when personal information rights are infringed or there is significant infringement risk, effectively preventing potential GenAI risks and curbing damage signs.

Conclusion

This study has explored the contemporary inheritance of Lingnan Aoyu Dance within the framework of the digital intelligence era. The research affirms that digital technologies offer transformative potentials for this intangible cultural heritage form by enabling the transition from offline-only transmission to integrated online-offline platforms, broadening participation from a limited circle of inheritors to a multi-subject collaborative network, facilitating the modernization of performance content while preserving its core symbolism, and enhancing the precision and efficiency of its dissemination through data-driven methods.

However, the investigation also uncovers significant hurdles that hinder the full realization of these potentials. These include a predominant focus on technological form over cultural depth, a critical gap in the succession of inheritors coupled with their generally low digital literacy, a tendency towards the homogenization and de-contextualization of cultural content in digital adaptations, and a lag in the systematic digital preservation of traditional resources.

To address these challenges and harness opportunities, a comprehensive and strategic approach is necessary. The proposed optimization strategy emphasizes several interconnected pathways: First, a foundational effort in digital empowerment for resource protection, involving the creation of detailed digital assets and intelligent management systems. Second, the construction of a robust "multi-agent collaboration" mechanism that enhances the digital capacity of existing inheritors, actively engages and cultivates youth participation, and fosters cross-sector innovation among government, industry, and academia. Third, the establishment of a "systematic guarantee" support system encompassing targeted policies, financial mechanisms, and integrated digital inheritance platforms. Fourth, the strategic use of intelligent technologies to create rich, immersive, and accessible cultural experiences that resonate with contemporary audiences both domestically and internationally. Finally, the proactive exploration of "international communication" paths that adapt content for global audiences while steadfastly maintaining the unique "Lingnan flavor".

In summary, the successful contemporary inheritance of Lingnan Aoyu Dance in the digital intelligence era is not a matter of simple technological adoption. It is a complex, holistic process that must strategically balance innovation with authenticity, technological capability with cultural meaning, and broad accessibility with dedicated, deep transmission. By anchoring itself in cultural connotation, using digital intelligence as a supportive tool rather than an end, and mobilizing the coordinated efforts of diverse societal forces, Lingnan Aoyu Dance can navigate the challenges of the digital age. This approach will not only ensure its vital continuity as a living tradition but also allow it to achieve renewed vitality and relevance, contributing to the dynamic preservation and innovative evolution of Lingnan's intangible cultural heritage dance landscape. Future research could longitudinally track the implementation of such strategies and delve deeper into the specific pedagogical and aesthetic impacts of individual technologies like AI and VR on intangible dance inheritance.

Funding

This work was not supported by any funds.

Acknowledgements

The author would like to show sincere thanks to those techniques who have contributed to this research.

Conflicts of Interest

The author declares no conflict of interest.

References

[1] Deng, Z., Xiang, H., Tang, W., Cheng, H., Qin, Q. (2024) BP neural network-enhanced system for employment and mental health support for college students. *International Journal of Information and Communication Technology Education (IJICTE)*, 20(1), 1-19.

[2] Twinomurinzi, H., Gumbo, S. (2025) Generative AI: concerns, usage, challenges, opportunities and

sentiments. *South African Computer Journal*, 37(1), 120-145.

[3] Mokyr, J. (2018) The past and the future of innovation: some lessons from economic history. *Explorations in Economic History*, 69, 13-26.

[4] Cano-Marin, E. (2024) The transformative potential of Generative Artificial Intelligence (GenAI) in business: a text mining analysis on innovation data sources. *ESIC Market*, 55(2), e333-e333.

[5] Christodorescu, M., Craven, R., Feizi, S., Gong, N., Hoffmann, M., Jha, S., Turek, M. (2024) Securing the future of GenAI: policy and technology. *arXiv preprint arXiv:2407.*12999.

[6] Arcila, B. B. (2023) Is it a platform? Is it a search engine? It's chatgpt! the european liability regime for large language models. *J. Free Speech L.*, 3, 455.

[7] Deng, Z., Ma, W., Han, Q. L., Zhou, W., Zhu, X., Wen, S., Xiang, Y. (2025) Exploring DeepSeek: a survey on advances, applications, challenges and future directions. *IEEE/CAA Journal of Automatica Sinica*, 12(5), 872-893.

[8] Cai, Y., Tian, S. (2025) Student translators' web-based vs. GenAI-based information-seeking behavior in translation process: a comparative study. *Education and Information Technologies*, 1-29.

[9] Izuchukwu, A. (2025) Data privacy rights and regulatory challenges of Generative AI models: a case study approach. *NEL Rev.*, 11, 16.

[10] Ferrara, E. (2024) GenAI against humanity: Nefarious applications of generative artificial intelligence and large language models. *Journal of Computational Social Science*, 7(1), 549-569.

[11] Wu, X., Duan, R., Ni, J. (2024) Unveiling security, privacy, and ethical concerns of ChatGPT. *Journal of Information and Intelligence*, 2(2), 102-115.

[12] Laine, J., Minkkinen, M., Mäntymäki, M. (2025) Understanding the ethics of generative AI: established and new ethical principles. *Communications of the Association for Information Systems*, 56(1), 7.

[13] Fontana, A. G. (2025) Web scraping: Jurisprudence and legal doctrines. *The Journal of World Intellectual Property*, 28(1), 197-212.

[14] Yu, Z., Xu, Z., Qi, J. (2025) Disability-oriented data protection in AI-enabled assistive technologies: Bridging gaps in China's legal framework. *Disability and Rehabilitation: Assistive Technology*, 1-25.

[15] Atkinson, D., Morrison, J. (2024) A legal risk taxonomy for generative artificial intelligence. *arXiv preprint arXiv:2404.*09479.

[16] Liu, C., Zhu, T., Zhang, J., Zhou, W. (2022) Privacy intelligence: a survey on image privacy in online social networks. *ACM Computing Surveys*, 55(8), 1-35.