Dilemmas and Institutional Optimization of Artificial Intelligence Training Data Supply

Jinze Sang[1]

[1]Institute of Data Rule of Law, China University of Political Science and Law, Beijing, China. Email: 2401421143@cupl.edu.cn

Abstract: High-quality training data is critical to the rapid development of artificial intelligence (AI). However, the acquisition of such data is currently hampered by structural obstacles, the conflict between intellectual property rights and data accessibility, and concerns over security and privacy. Drawing on an analysis of legal practices both in China and abroad, this paper identifies the primary bottlenecks as the unequal distribution of interests and inadequate institutional alignment. To resolve these issues, the study proposes a structured framework designed to ensure a high-quality data supply. This framework focuses on refining data property rights, diversifying incentive mechanisms, and strengthening data-sharing platforms, with the ultimate goals of balancing interests, ensuring security, and fostering innovation. By combining these measures with collaborative governance, the paper presents a systematic legal solution to the challenges of AI data supply, supporting the technology's sustainable growth.

Keywords: Artificial Intelligence; Data Governance; Training Data; High-quality Data Supply System

Introduction

In recent years, the development of AI technology has witnessed unprecedented evolution. From ChatGPT to Sora, the functions of products have been significantly upgraded within just two years, fully demonstrating the rapid momentum of AI technological innovation. As an emerging discipline, AI has continuously made breakthroughs in algorithms, data, computing power and other aspects since its birth in the 1950s. Currently, it has become one of the most revolutionary technologies, not only showing great potential in fields such as automation and machine learning, but also being widely applied in various industries including medical care, finance, and manufacturing, driving human society towards a more efficient and intelligent direction. With the continuous progress of technology, AI plays an increasingly significant role in solving complex problems and improving work efficiency. Its rapid development has also prompted countries around the world to accelerate the formulation of relevant industrial policies to seize the strategic high ground of scientific and technological development.

Data is a fundamental factor of production for digitalization, networking, and intellectualization. Enterprises rely on massive, timely, and high-quality data factors to match the needs of all parties through data relevance, activate data value, and promote the quality and efficiency improvement of enterprise products and services. High-quality training data plays a crucial role in the development of AI. For AI products, high-quality training data is the key to improving functions. Rich and diverse datasets enable models to accurately recognize images and texts, thereby boosting recognition rates and user experience. Specifically, large model pre-training depends on large-scale, high-quality, and diverse datasets to understand complex problems and improve generalization capabilities across various application scenarios. Without the support of such data, achieving genuine breakthroughs in AI technology would be difficult.

Although training data is crucial for the development of AI technology, its supply process faces many legal issues that restrict the development of the AI industry. Firstly, the legal boundary of data acquisition is vague, especially in scenarios such as web crawling and cross-platform data sharing. There is still a lack of clear standards for defining legal data acquisition behaviors and illegal intrusion or data theft behaviors[1]. Secondly, issues such as the use of copyrighted works in training data, the circulation of personal information, and enterprise data sharing have also triggered disputes in terms of intellectual property protection, privacy security, and technical ethics[2]. Generative AI may face risks such as compliance, data breach, and bias during the process of training data collection and processing. These issues not only threaten the healthy development of technology, but also may disrupt social order. In terms of training data supply, on the one hand, the current laws have relatively principled provisions on the training data review system, lacking specific implementation guidelines, which makes it difficult for enterprises to effectively eliminate biased data during the data preprocessing and model training process. On the other hand, the ambiguity of public data authorization and operation rules further hinders the process of data opening, especially in terms of the definition of public data ownership and profit distribution mechanism, which urgently requires more clear legal norms[3]. The existence of these problems not only increases the difficulty of data governance work, but also poses severe challenges to the sustainable development of the AI industry[1].

Consequently, existing research and practices reveal several critical gaps. Firstly, regarding the refined standards for the legality of training data acquisition, current studies predominantly focus on macro-level principles, lacking operational guidelines for specific scenarios. For example, in the context of data crawling, although existing literature has proposed reference factors, a unified standard has not yet been formed, leading to difficulties in achieving consistency between law enforcement and judicial institutions in practice[4]. Moreover, with the application of emerging technologies such as knowledge distillation, the boundary of training data use has become increasingly blurred. Protecting the legitimate rights of data owners while promoting technological innovation remains an urgent problem to be addressed[5]. Secondly, concerning legal regulation under emerging technologies, existing studies have not fully accounted for the complexity introduced by the rapid iteration of AI. The large-scale training of generative AI models relies on massive data involving complex ownership relationships and sensitive information; however, how to promote data circulation while ensuring security remains insufficiently addressed in current

research[6]. Simultaneously, regarding enterprise data sharing, existing literature mostly remains at the theoretical level, lacking in-depth analysis of practical cases and solutions. Thirdly, in terms of international cooperation and legal coordination, current studies have paid insufficient attention to legal conflicts arising from cross-border data flows. As the cross-border acquisition and use of training data become prevalent alongside the global development of AI, significant disparities in data protection standards and intellectual property rules persist across regions. Therefore, building a unified data governance framework remains a critical direction for future research.

1. Practical challenges of training data supply

Despite the critical role of high-quality training data as a fundamental production factor in the artificial intelligence ecosystem, its actual supply is currently impeded by a complex array of practical barriers. These challenges are not merely technical but deeply rooted in the socio-economic structure of data ownership, the legal frameworks governing intellectual property, and the ethical imperatives of security. This section dissects the dilemmas of training data supply from three key dimensions: the structural obstacles arising from data monopolies and fragmentation, the inherent conflict between copyright protection and data sharing, and the escalating risks regarding data security and privacy compliance.

1.1 Structural obstacles to data acquisition from monopolies and fragmentation

There are numerous difficulties in the process of AI training data acquisition. Limited data sources are the primary problem. Although we are in an era of data explosion with massive amounts of data generated every day, high-quality data suitable for AI training is extremely scarce. Taking the humanoid robot scenario as an example, the environment it faces is diverse and complex, and the required data is difficult to collect, resulting in a serious shortage of data supply. Based on the theoretical framework of Informational Capitalism[7] (also known as Surveillance Capitalism or Data Capitalism), from the perspective of ownership of means of production, there are structural obstacles to data acquisition. The penetration rate of five major Internet platforms such as Douyin among 238 million high-value active users across the network reaches 85.7%. Super-large platforms rely on their extensive business layout and large user groups to accumulate a large amount of original data resources such as system transaction data and user behavior data[8]. Super-large platforms continue to focus on algorithm competition and talent competition, carry out killer acquisitions around algorithm talents[9], dominate the iteration of cutting-edge algorithms such as GAN and VAE, and consolidate their monopoly position in synthetic data production through technical patents and core code closure.

Data dispersion is also a major obstacle to data acquisition. Data is scattered in different institutions, departments, and platforms, and has different formats. For example, policy data released by the government has scattered release channels and inconsistent formats. Extracting effective information requires a lot of human and material resources, and it is also prone to cause information update lag. It is even more difficult for enterprises to screen out data relevant to themselves from massive information. The phenomenon of data silos is widespread. All parties are reluctant to share data due to interest considerations, which further limits the channels for data

acquisition. These problems have seriously hindered the acquisition of AI training data and affected the development process of AI technology.

1.2 Conflict between intellectual property protection and data sharing efficiency

Existing theoretical studies not only focus on the antitrust regulation of specific monopolistic behaviors such as traditional data-driven algorithmic collusion[10], platform most-favored-nation clauses[11], algorithmic price discrimination[12], and platform self-preferencing[13], but also discuss the risk governance rules of generative AI and data risk governance rules[14]. However, there are still sharp contradictions between the use of AI training data and intellectual property protection. On the one hand, AI technological innovation requires massive data support for the optimization and upgrading of training algorithm models; on the other hand, data often contains a large number of works protected by intellectual property rights.

There is controversy over whether behaviors such as copying training data in large model training infringe copyright. More and more lawsuits between AI enterprises and copyright owners have emerged in the United States. In China, there are also cases where copyright owners sue AI painting software companies for using their works to train models without permission. These cases reflect the conflict between the use of training data and copyright protection. In terms of data sharing, intellectual property protection restricts the free circulation of data. In order to protect intellectual property rights, data owners may set many restrictions on the use of data, making it difficult for data to be widely shared for AI training. However, the development of AI technology urgently needs the sharing of a large amount of data to promote technological innovation. This contradiction makes it an urgent problem to be solved to balance intellectual property protection and data sharing while promoting the development of AI.

1.3 Data security risks and challenges in privacy protection

AI training data faces many security and privacy risks during the process of collection, storage, and use. In the data collection stage, the data sources are complex, making it difficult to ensure the authenticity and security of the data. Data collection methods such as web crawlers may obtain forged or tampered data, leading to biases in the trained models, and even potential malicious use. According to the Personal Information Protection Law and other relevant laws and regulations, the collection and use of personal information must follow the principles of legality, legitimacy, and necessity, and obtain the explicit consent of the information subject[15]. However, in the training process of generative AI, the collection and processing of large-scale data often involves the use of massive personal information, which poses a severe challenge to the existing legal framework. ChatGPT used a large amount of dialogue data containing personal information during its training process. Whether the use of these data complies with the principle of informed consent and how to realize the effective use of data while ensuring privacy have become urgent problems to be solved[16].

For data storage, Article 9 of the “Interim Measures for the Management of Generative Artificial Intelligence Services” requires service providers to take necessary measures to prevent data breach and abuse, but the specific implementation rules still need to be refined[17]. Storage devices may face the risk of physical damage, leading to data loss or breach. If there are vulnerabilities in the data management system, hackers may take the opportunity to invade, steal or destroy data.

During the data use process, privacy leakage problems may occur in both the model training and prediction stages. Untrusted servers or participants can use the intermediate results of training to construct attack models to infringe on user privacy. Attackers can also steal target training data by accessing the model prediction interface. These risks will not only cause serious harm to personal privacy, but also affect the trade secrets of enterprises, and even threaten national security. Data security and privacy protection issues have become a major obstacle on the path of AI development, which must be highly valued and effectively addressed.

2. Policy practices and limitations on the supply of AI training data

In terms of legislation on AI training data supply, the United States follows the traditional legislation-first approach. In order to ensure its global leadership in the field of AI, it has made many attempts at the administrative order and legislative levels in recent years. Although some bills have not yet become laws, the proposal and debate process of relevant bills provide important reference for the subsequent formulation and implementation of mature AI security governance regulations. The "California Consumer Privacy Act" (CCPA) endows consumers with the right to know, delete, and opt out of their personal data, which puts forward clear requirements for the processing of personal information involved in training data[18]. In the case of HiQ v. LinkedIn, there was a dispute over whether the act of third-party crawling of public data from social media platforms constitutes unfair competition. The court finally ruled that LinkedIn could not prevent HiQ from crawling its public data. This judgment provided an important reference for the legal boundary of data crawling.

The European Union regulates the supply of training data through a series of data bills. The General Data Protection Regulation (GDPR) in 2016 had a profound impact on data protection in various member states. The ePrivacy Regulation (draft) released in 2021 supplemented the GDPR in the field of electronic communications. The European Data Strategy and European Artificial Intelligence White Paper released in 2020 outlined the blueprint for building a healthy common data space. The European Union also intends to establish a global standard for AI supervision through the Artificial Intelligence Act (AI Act), which to a certain extent avoids potential harms such as discrimination and surveillance, and clarifies control measures to reduce risks. Article 6 of the GDPR stipulates six legal bases for data processing, such as consent, performance of contracts, and public interest. These provisions directly affect the compliance of generative AI in the process of training data acquisition[19]. In addition, the European Union also put forward additional requirements on the data use of high-risk AI systems through the draft AI Act , such as prohibiting data analysis based on sensitive personal characteristics, which further limits the sources and scope of use of training data[20]. In the case of Google v. CNIL, Google was fined heavily for failing to comply with the requirements of the "right to be forgotten" in the GDPR, which shows that the European Union has extremely strict law enforcement in data protection[19].

China attaches great importance to the supply of AI training data and has issued a series of relevant policies. Normative documents such as the Development Plan for the Next Generation Artificial Intelligence have repeatedly mentioned industrial

policies such as building public data resource libraries, standard test data sets, and cloud service platforms for AI, so as to promote the orderly opening of public data by classification and grade and expand high-quality public training data resources. At the specific implementation level, the State Intellectual Property Office has clearly listed stem cells and regenerative medicine, immunotherapy, cell therapy, etc. as one of the key industries that the country focuses on developing and urgently needs intellectual property support. The China National Center for Biotechnology Development has also announced the proposed projects of the key special projects of the national key R&D plan for 2018, such as "Stem Cell and Transformation Research". These policies provide support for the supply of AI training data from different angles and create a good development environment[21]. The "Interim Measures for the Management of Generative Artificial Intelligence Services" puts forward specific requirements on the source legality, quality requirements, and labeling specifications of training data, reflecting China's refined governance ideas in the field of AI[17]. On the one hand, courts have gradually clarified the legal boundary of training data acquisition through judgments in specific cases, providing relatively clear compliance guidelines for enterprises; on the other hand, the existing legal system is still insufficient in dealing with complex problems brought by emerging technologies. For example, issues such as the interest distribution mechanism of data sharing and the intellectual property ownership of knowledge distillation technology have not been fully resolved[16]. Therefore, it is necessary to further improve the relevant legal system in the future to adapt to the rapid development of AI technology.

3. Core institutional construction for guaranteeing the high-quality supply of training data

To effectively resolve the structural dilemmas and interest conflicts impeding the supply of AI training data, it is not sufficient to rely solely on market forces; a systematic legal and institutional framework is required. This section proposes a comprehensive guarantee system centered on three strategic pillars: property rights, incentives, and platform infrastructure. Specifically, it advocates for a refined definition of data property rights to clarify legal boundaries, the coordinated allocation of diversified incentive mechanisms to mobilize high-quality resources, and the robust institutional support for data sharing platforms to ensure secure circulation. These measures aim to reconstruct the interest distribution pattern among stakeholders, thereby fostering a sustainable environment for AI innovation.

3.1 Refined design of data property rights system and intellectual property rules

In the institutional construction for the high-quality supply of AI training data, the data property rights system is crucial. The ownership of data property rights needs to be clarified. Currently, training data comes from a variety of sources, involving multiple subjects such as individuals and enterprises. If the ownership is not clear, disputes are likely to arise. For example, there is a dispute over the ownership of information posted by individuals on social platforms if it is used for AI training. To solve this problem, we can learn from the "Twenty Provisions on Data" and establish a property rights operation mechanism that separates data resource ownership, data processing and use rights, and data product operation rights.

Constructing intellectual property rules adapted to the needs of AI training data is an important part of improving the legal system guarantee. With the development of

generative AI technology, copyrighted works are increasingly used in training data, but there are still great disputes over how to reasonably define the scope of their use. Therefore, it is recommended to start from two aspects: first, clarify the fair use boundary of copyrighted works in training data. For example, through judicial interpretation or legislative revision, include specific types of AI training into the scope of "fair use"; second, improve the intellectual property authorization mechanism and explore flexible and diverse licensing models to reduce the cost of enterprises in obtaining training data and improve efficiency[6]. In addition, considering the challenges brought by emerging technologies such as knowledge distillation to the ownership of intellectual property rights, it is also necessary to study the contribution and right distribution of different participants in the process of knowledge generation, so as to form a more fair and reasonable intellectual property protection system[17].

In terms of interest protection, a sound mechanism should be constructed. When using data, training data processors need to respect the prior rights of data source owners, such as copyright and portrait rights. For public data, the scope of opening and use conditions should be clarified to prevent abuse. For enterprise data, it is necessary to protect the trade secrets of enterprises and promote the reasonable circulation of data. Through the formulation of detailed rules, the fair and reasonable distribution of interests in the process of data use is ensured, so that data subjects can fully enjoy the benefits brought by data, effectively solve the contradiction between data use and intellectual property rights, and provide a solid property rights guarantee for the high-quality supply of AI training data. Break through the lag limitation of traditional laws, take the technology development trend as the guide, and formulate special legislation to clarify the core rules of training data supply. We can learn from the refined governance ideas of the European Union's GDPR, combine the data property rights separation principle established in China's "Twenty Provisions on Data", and clarify the boundaries of data resource ownership, processing and use rights, and product operation rights in special laws, and refine the legality standards of data use in technical scenarios such as knowledge distillation. At the same time, dynamically fill the rule gaps through judicial interpretations and guiding cases. For example, regarding the scope of "fair use" of copyrighted works in generative AI training, clarify specific applicable conditions such as "non-commercial training" and "small proportion quotation", so as to not only ensure the space for technological innovation, but also prevent the risk of abuse of rights.

3.2 Coordinated allocation of diversified incentive mechanisms and benefit distribution

In order to promote data providers to actively participate in the supply of AI training data, it is necessary to carefully design incentive mechanisms. Material incentives are important means. A data trading market can be established, and data pricing rules can be clarified to allow data providers to obtain benefits by selling data. Data pricing should comprehensively consider factors such as data quality, scarcity, and application value. Methods such as pay-per-use and time-based charging can be adopted to ensure that data providers can obtain reasonable returns according to the value of data. In the field of data trading, use blockchain technology to build a trusted deposit platform to realize the whole-chain traceability of training data sources, processing, and circulation, and solve the problems of ownership definition and transaction security; in the supervision link, introduce AI technology to develop a compliance review system, and automatically implement requirements such as data

desensitization and permission management through smart contracts to improve the real-time supervision efficiency of behaviors such as data crawling and cross-border transmission. At the same time, it is necessary to clarify the boundaries of technology application in laws, such as stipulating the judicial effect of blockchain deposit data, and preventing new risks caused by technology abuse.

In order to promote enterprise data sharing, it is necessary to clarify the interest distribution and risk-bearing rules in the sharing process through legal means. First of all, legislation should be adopted to establish the basic principles of data sharing, such as fairness, transparency, and mutual benefit, to ensure that enterprises participating in sharing can achieve a win-win situation in cooperation[4]. Secondly, it is necessary to formulate detailed operating specifications, clarify the scope, methods, and duration of data sharing, and establish a corresponding supervision mechanism to prevent data abuse or improper use[19]. In addition, regarding data security issues, it is recommended to introduce advanced technologies such as blockchain to realize the secure sharing and controllable access of data through distributed ledgers and smart contracts. Blockchain technology can effectively record the whole process of data sharing, ensure the authenticity and immutability of data, and thus enhance enterprises' trust in data sharing[23]. It is also possible to establish an integral incentive and data contribution integral system. The integrals can be used to exchange data services, technical support, etc., so as to enhance the participation enthusiasm of data providers. Incentive measures should be combined with data security and privacy protection to ensure that data providers do not face security and privacy risks while providing data. Through a sound incentive mechanism, the enthusiasm of data providers is fully mobilized to provide a continuous driving force for the supply of AI training data.

3.3 Regulatory support and architectural optimization for data sharing platforms

Data sharing platforms are the key to realizing the effective circulation and sharing of AI training data. The platform architecture design should be scientific and reasonable. Technologies such as cloud computing and big data should be adopted to build a unified cloud platform overall architecture. The platform can be divided into a five-layer data exchange ecosystem including data layer, consent layer, data supply layer, exchange layer, and consumption layer to ensure the whole process of data from collection, storage, processing to use is safe and controllable. Since training data involves complex legal issues, such as the definition of data ownership, the identification of intellectual property infringement, and the division of personal information protection responsibilities, the judgment results of courts in different regions in similar cases often vary greatly. This not only weakens the credibility of the judiciary, but also increases the uncertainty of enterprise operations[31]. Therefore, it is recommended to unify the judicial judgment standards by issuing judicial interpretations or releasing guiding cases, and clarify the handling principles and methods for various controversial issues. In data crawling cases, the identification standard of "substantial damage" should be clarified; in disputes over the use of copyrighted works, the specific applicable conditions of "fair use" need to be refined[32].

Establish a national AI ethics committee, which includes technical experts, legal scholars, public representatives and other parties, to evaluate the ethical risks in the supply of training data and put forward legislative suggestions; encourage industry

associations to formulate self-regulatory norms, such as the qualification review standards of data trading platforms and the security guidelines for enterprise data sharing; smooth the channels for public participation, and ensure that legal rules take into account both technological innovation and people's well-being through legislative hearings, online opinion collection and other methods. In addition, promote the establishment of a normalized communication mechanism among enterprises, scientific research institutions, and judicial organs to timely feedback new problems in practice. In terms of functions, the platform should have preprocessing functions such as data cleaning, labeling, and integration to ensure data quality. It should also provide convenient services such as data retrieval, download, and analysis to facilitate users to quickly obtain the required data. The platform should establish a sound data security and privacy protection mechanism, conduct encrypted storage and access control of data, and prevent data breach and abuse. Through the construction of a fully functional and secure data sharing platform, data silos are broken, cross-departmental and cross-industry circulation and sharing of data are realized, rich high-quality data resources are provided for AI training, and the innovative development of AI technology is promoted.

4. Conclusion

The sustainable development of artificial intelligence is fundamentally contingent upon the effective supply of high-quality training data. However, current practices are severely constrained by multiple dilemmas, including structural obstacles to data acquisition, acute conflicts between intellectual property protection and data sharing, and pervasive risks regarding data security and privacy. While domestic and international legal policies have begun to address these issues, existing frameworks still suffer from limitations such as imbalanced interest distribution, vague implementation standards, and a lack of regulatory coordination.

To overcome these challenges, this study proposes the construction of a comprehensive guarantee system for the high-quality supply of training data, anchored in the goals of interest balance, security controllability, and innovation adaptation. This system relies on three core institutional pillars: first, the refined design of the data property rights system to clarify ownership and usage boundaries; second, the coordinated allocation of diversified incentive mechanisms to mobilize high-quality data resources; and third, the regulatory and technical support for data sharing platforms to ensure secure circulation. By integrating these institutional innovations with multi-subject collaborative governance and cross-domain legal coordination, a systematic legal solution can be formed to resolve the dilemma of data supply, thereby providing a robust foundation for the continuous innovation and application of AI technology.

Acknowledgments

[1] Zhao Jingwu, Zhou Ruijue. On the Transformation of Digital Law Research Paradigm: Risk Systematic Governance[J]. Qiushi Journal, 2024(4):136-147.
[2] Chen Hongguang, An Shifeng. Risks and Governance of Generative Artificial Intelligence Training Data[J]. Journal of Hainan Open University, 2024(1):110-118.
[3] Zhang Linghan. Accelerating the Construction of Chinese Training Data Corpus for Artificial Intelligence Large Models[J]. Academic Front, 2024(13):57-71.

[4] Zhao Jingwu. On the Concept and Research Orientation of Digital Law - Also on What Kind of Artificial Intelligence Law We Need[J]. Journal of East China University of Political Science and Law, 2024(4):64-75.
[5] Zhang Ping. Breakthrough Innovation of Generative Artificial Intelligence Requires Good Law and Good Governance - Taking the Legality of Data Training as an Example[J]. New Economy Herald, 2023(8):26-28.
[6] Chen Bing. Risk Challenges Brought by the Innovative Development of General Artificial Intelligence and Its Legal Response[J]. Intellectual Property, 2023(8):53-73.
[7] Castells M. The Information Age: Economy, Society and Culture, The Rise of the Network Society[M]. 1996:14-16; Ascárate R J. Dan Schiller: How to Think about Information[J]. MEDIENwissenschaft: Rezensionen|Reviews, 2007:373-375.
[8] See QuestMobile official website, https://www.questmobile.com.cn/research/report/1896846900944015361, accessed on May 17, 2025.
[9] Barnett, Jonathan. The Case Against Preemptive Antitrust in the Generative Artificial Intelligence Ecosystem[J]. Artificial Intelligence and Competition Policy, Concurrences (2024), USC CLASS Research Paper 2419 (2024).
[10] Ezrachi A, Stucke M E. Artificial Intelligence & Collusion: When Computers Inhibit Competition[J]. U. Ill. L. Rev., 2017:1775; Han Wei (Ed.). Research on Digital Market Competition Policy[M]. Law Press, 2017:344-345; Shi Jianzhong. Research on the Extended Application of the Collective Dominance System to Algorithmic Tacit Collusion[J]. China Legal Science, 2020(2):100-101; Wan Jiang. Digital Economy and Antitrust Law: From the Perspective of Theory, Practice and International Comparison[M]. Law Press, 2022:131.
[11] Smith, W. Stephen. When Most-Favored Is Disfavored: A Counselor's Guide to MFNs[J]. Antitrust, 2012:10; Tan Chen. Antitrust Regulation of Most-Favored-Nation Clauses in the Internet Platform Economy[J]. Journal of Shanghai University of Finance and Economics, 2020(2):141+148-149.
[12] Ye Ming, Guo Jianglan. Legal Regulation of Algorithmic Price Discrimination in the Digital Economy Era[J]. Price Monthly, 2020(3):33-40; Wan Jiang. Digital Economy and Antitrust Law: From the Perspective of Theory, Practice and International Comparison[M]. Law Press, 2022:153-154; Supreme People's Court (2021) Zui Gao Fa Zhi Min Zhong No. 1020.
[13] Chen Yongwei. A Review of the U.S. House of Representatives' "Investigation Report on the State of Digital Market Competition"[J]. Competition Policy Research, 2020(5).
[14] Liu Yanhong. Three Major Security Risks of Generative Artificial Intelligence and Their Legal Regulation - Taking ChatGPT as an Example[J]. East China Law Review, 2023(4):29-43; Zhang Linghan, Yu Lin. From Traditional Governance to Agile Governance: Innovation of Governance Paradigm for Generative Artificial Intelligence[J]. E-Government, 2023(9):2-13.
[15] Liu Yanhong. Algorithm-Centric Governance Model for the Judicial Security Risks of Artificial Intelligence[J]. East China Law Review, 2024(4):83-94.
[16] Shan Yong, Wang Yi. Knowledge Fraud from "World 3": Criminal Risk Response to Generative Artificial Intelligence[J]. Journal of Southwest University of Political Science and Law, 2023(4):70-85.
[17] Wu Jing. Data Risks of Generative Artificial Intelligence and Their Legal Regulation - Taking ChatGPT as an Example[J]. Science and Technology Management Research, 2024(5):192-198.

[18] Deng Zhenyu. Challenges and Response Paths for the Responsible Development of Generative Artificial Intelligence[J]. Cybersecurity and Data Governance, 2024(7):69-76.
[19] Xing Luyuan, Shen Xinyi, Wang Jiayi. Research on the Regulation Path of Generative Artificial Intelligence Training Data Risks[J]. Cybersecurity and Data Governance, 2024(1):10-18.
[20] Song Huajian. On the Legal Risks and Governance Paths of Generative Artificial Intelligence[J]. Journal of Beijing Institute of Technology (Social Sciences Edition), 2024(3):134-143.
[21] Zhao Jingwu. On the Institutional Construction of High-Quality Supply of Artificial Intelligence Training Data[J]. China Legal Review, 2025(1):92-104.
[22] Hu Anjun. Several Key Points for Promoting the In-depth Development of Artificial Intelligence[J]. China Development Observation, 2020(17):61-65.