

Data Audit Rules for Generative AI: Strategic Design and Governance Pathways

Song Lang

(Lanzhou Vocational Technical College, Lanzhou 730070, China)

Abstract: The development of generative artificial intelligence (AI) models, typically represented by ChatGPT, has become a new wave in the growth of the science and technology industry. This emerging new technology has aroused worldwide concern about its security risks, especially legal risks. Among these, data security risk is a new type of risk brought about by large generative AI models. Because of its market-oriented, routine and professional advantages, data auditing has become a feasible option for addressing data security risks, yet China's institutional and legal supply in the field of data auditing remains insufficient. China should focus as early as possible on building and improving the industrial policies and the industrial legal basis for data auditing, and should shift the emphasis of building the data audit rule system from a one-sided protection of the value of security to an all-dimensional protection of data subjects. While improving domestic legal norms, China should also respond at the legislative level to challenges worldwide, and should devote itself to developing data audit norms that can serve global data flows and trade, so as to occupy a favorable position in the development of data audit norms at the level of international law.

Keywords: generative artificial intelligence; data risk; data audit; digital trade

1 Statement of the Problem

In November 2022, the American company OpenAI released a new-generation generative AI language model named ChatGPT. Within less than two months, its user base reached 100 million, surpassing the nine-month fastest-growth record previously held by TikTok. ChatGPT's powerful language generation capability attracted wide attention from such industries as scientific research, education, the media and computing, and triggered a boom in the development of various related application fields. At the user's instruction, ChatGPT can autonomously compose academic papers, literary works, work documents, programming code and so on. According to users' experience, the content it generates has already reached a fairly high professional standard. In the global AI wave driven by ChatGPT, the marked improvement in productivity brought about by the continuous iteration of technology is clearly visible, and voices proclaiming the "arrival of the singularity" can be heard without end. Yet this has also aroused worldwide concern about the various risks brought by the development of AI.

Data security risk is one of the main types of risk brought by generative AI technology. As a natural language processing model, ChatGPT requires a huge amount of data and computing power for training and operation. This means that it needs to collect and process large quantities of user data. When users interact with ChatGPT, it collects and stores information related to the interaction, including users' input, output, language, timestamps and device information. Such data can be used to train and improve ChatGPT and, at the same time, to raise service quality so as to meet users' needs. In addition, OpenAI may potentially use these data for other purposes, such as research or commercial objectives. Therefore, user privacy and data protection are issues of vital importance in the legal and ethical aspects concerning ChatGPT. The main manifestation of data security risk is a security threat. Security is one of the basic values pursued by law; here, a security threat refers to a potential risk that may lead to illegal or criminal activities and may cause major harm to individual or public interests. It mainly includes three types: the security risk of training data, the security risk of generated data, and the security risk of data privacy.

1.1 Security Risk of Training Data

The security risk of training data refers to the possibility that the data used by ChatGPT may contain false or illegal information. Although OpenAI has endeavored to avoid outputting content involving discrimination, bias, attacks, pornography, violence and bloodshed, the opacity of ChatGPT's training process makes it impossible to ensure that such problematic data are accurately filtered out; as a result, these data will affect the content generated by ChatGPT, which may carry an unlawful tendency. ChatGPT states that it will abide by legal and ethical norms, but it cannot fully identify all illegal information. If illegal information is found, it should be reported to a legal institution or an authoritative department.

ChatGPT's training data come from the pre-training stage and from user interactions. The data of the pre-training stage may contain illegal content. Data from user interactions are also used for training, but malicious users may, through cleverly designed questions, feed illegal information into ChatGPT, thereby increasing the risk. For example, a user may ask how to avoid areas of the pornography industry, and ChatGPT may provide false information. This shows that ChatGPT cannot completely filter out false or illegal information.

1.2 Security Risk of Generated Data

ChatGPT can generate false or illegal information, and defects in the training data or imperfections in the algorithm are both possible causes. It is able to fabricate false information, for example by inventing references when writing a paper. Some users have also reported that ChatGPT fabricates untrue news information, becoming a source of social panic and rumors and very likely causing serious harm to society. Tests show that ChatGPT can quickly adapt information to create convincing but unsourced false information. There is currently no effective solution to this problem, so ChatGPT may become a powerful tool for spreading misinformation and may cause false information to spread more widely.

Besides being able to fabricate false information, ChatGPT may also generate information involving illegality and even crime. When ChatGPT was first released, an engineer used a method of "luring" to induce ChatGPT to generate a "plan for destroying humanity". The plan included steps such as intruding into the computer systems of various countries, taking control of communications and disrupting transportation systems, and even the corresponding Python code. The threat intelligence company Recorded Future found more than 1,500 references on the dark web related to the use of ChatGPT for malware development. Some media reports also pointed out that after multiple rounds of dialogue, ChatGPT may become irritable and may even begin to propose criminal ideas. Although ChatGPT lacks self-awareness, its capacity to generate illegal or criminal content is mainly exploited by malicious individuals as a tool, and it has already significantly lowered the threshold for many high-tech crimes.

1.3 Security Risk of Data Privacy

Besides the two types of data security risk mentioned above, the security risk of data privacy is also a key issue. ChatGPT requires data provided through user interactions for training, but because its algorithm is opaque, how it handles users' private data cannot be verified. Unlike many domestic applications, which collect detailed personal profiles and set privacy terms at registration, ChatGPT has no such steps. Although OpenAI states that users' data will be used for training only after personal information has been deleted, it remains doubtful whether the private information that users may disclose during interactions can be completely filtered out so as to protect their data. Some users have reported that their ChatGPT conversations appeared in other users' conversation lists. OpenAI has also admitted that the ChatGPT Plus service may have leaked some users' privacy and payment information; although OpenAI promises not to expose complete credit card numbers, public concern about ChatGPT leaking private data persists. The Italian Data Protection Authority has banned the use of ChatGPT and launched an investigation into OpenAI, holding that ChatGPT violates the provisions of the European General Data Protection Regulation and that its large-scale collection and storage of personal data for training algorithms lacks legitimacy.

Apart from the risk of leaking personal private data, interactions with ChatGPT may also involve commercial and

state secrets. Such information is more sensitive than personal privacy, and its leakage may bring greater risks. For example, less than 20 days after Samsung Electronics introduced ChatGPT, an incident of confidential information leakage occurred, in which semiconductor equipment data and product information were stored in ChatGPT's learning database.

2 SWOT Analysis of the Strategy for Building Data Audit Rules

In the face of the new types of data security risk involved in a series of development and use processes of generative AI—algorithm development, algorithm application and generated results—the existing legal framework for data protection is indeed no longer applicable. However, in addressing this problem, the establishment of regulatory norms must likewise be exercised with a certain restraint: the legislative purpose of regulatory norms is to promote the sound use and development of AI, not the opposite. As the driving force behind the development of AI technology, data itself is not the object we need to regulate; rather, what is regulated is the conduct of collecting, processing, using and generating data, and the ultimate purpose is to promote the safer flow and use of data. The overall design and strategic choice for building a new regulatory rule system for the AI industry must be made with extreme caution. The volume of data involved in the AI industrial chain is unprecedented, and it is neither realistic nor feasible to make the administrative power of public authorities assume the task of specific supervision and management of the data application process. Whether it is the regulatory efficiency required by industrial development or the administrative costs that public authorities would have to bear as a result, both pose great problems in the face of an unprecedented scale of data. Therefore, data auditing—a regulatory model of a market-oriented nature—has become a good option, and establishing and improving sound legal norms for data auditing is an important legal countermeasure.

Data auditing refers to review activities carried out, in accordance with legal norms, technical standards and enterprises' internal control requirements, by professional auditors or regulatory agencies, and covering corporate data protection management and accountability, employee data protection training, personal data security, an overview of the protection of personal data rights, information sharing, records management, as well as data protection impact assessment and data risk management systems. The data audit system can comprehensively and objectively reflect the activities of data processors and evaluate their performance of the obligation of data security governance. In the course of a data audit, the audit team may not only enter the enterprise to make inquiries and conduct interviews, but may also access technical documents and consult relevant materials, use specific methods and tools to carry out such related assessment work as data leakage scanning, effectiveness testing of data desensitization and data quality testing, and may also issue security ratings, flag data security risks and put forward optimal improvement suggestions with respect to weak points and risk points.

Data auditing can run through every link of AI data utilization, and it usually includes the tracking and recording of data transmission, access and processing activities. This helps to ensure that the use of data complies with the rules and that problems can be traced and resolved once they occur. In the process of data transmission and processing, the integrity of data needs to be ensured and data must be prevented from being tampered with or damaged. Data auditing can help confirm whether any unauthorized changes have occurred to data during transmission. Data auditing may involve identity verification and access control for personnel who access data. This helps prevent unauthorized access and data leakage. Sometimes, in international data trade, the parties involved may choose to have a third party conduct the data audit so as to ensure the objectivity and neutrality of the audit process. Data auditing can be realized with the help of various technical tools, including logging, blockchain technology, encryption and security authentication. It can be said that data auditing is an important service industry for guaranteeing the sound development of the AI industry in the future. Only by establishing sound and complete data audit norms and promoting the development of the data auditing industry can the prevention of data security risks be institutionalized and made routine. In view of the governance effectiveness of the data audit system, China's regulatory authorities have already laid out a data

security audit system, under which a third-party professional institution is to audit the data security situation and issue an audit report at fixed intervals, so as to gradually establish a routine external supervision mechanism [4].

The SWOT model is a framework for analyzing the situation of the internal and external competitive environments and conditions closely related to the object of study [5]. The model arranges internal strengths and weaknesses together with external opportunities and threats in matrix form and, through comprehensive analysis, draws conclusions and assists decision-making; it is a strategic analysis model that has already been developed fairly maturely in academia. The use of the SWOT analysis method makes it possible to analyze fairly scientifically and clearly the current strategy for building data audit rules in the AI industry.

2.1 Internal Strengths (S): A Blue Ocean of Data, a Rapidly Growing Data Industry and Strong Policy Capacity

First, China, with a quarter of the world's population and the economic scale of the world's second largest economy, has an incomparable advantage in the scale of its data resources. China's data resources enjoy an advantage not only in magnitude; recent breakthroughs in China's chip industry have filled the gaps in its chip industrial chain and further consolidated China's position as the only country in the world with a complete range of industrial categories. This makes China's blue ocean of data incomparable not only in magnitude but also in complexity and richness, which surpass all others in the world. A blue ocean of data that is high in both quality and quantity provides abundant data samples and opportunities for the development of the data auditing industry. With such inherent advantages, the data auditing industry will not have to confine itself to supervising the process of data utilization; it can also deepen the analysis and processing of data in the course of auditing and promote the transformation of audit outputs into products. Data audit products can be used to analyze and evaluate the impact and effects of various government policies, commercial activities and social phenomena. Diversified operation of the data auditing industry becomes possible, and the room for the industry's growth is very broad.

Second, on the basis of the comparative advantage of its blue ocean of data, China's data industry has developed rapidly over the past few years, covering every link from data collection, storage and analysis to mining and application. The technologies and industrial chains in such fields as big data, AI and cloud computing have also continued to grow. The development of e-commerce has covered almost every link of China's economic activities, and both its industrial scale and industrial efficiency rank first in the world. The rapid development of these data-related industries and technologies provides technical support and infrastructure for data auditing, making it easier to carry out large-scale data analysis and audit work.

Finally, the Chinese government's administrative efficiency and policy capacity are very important internal strengths. That the policy capacity of the Chinese government is far higher than that of Western countries is a self-evident advantage. The report of the 20th National Congress of the Communist Party of China pointed out that China should step up efforts to build itself into a cyber power and a digital China. General Secretary Xi Jinping pointed out incisively that to accelerate the building of a digital China, it is necessary to adapt to the new historical orientation of China's development, fully implement the new development philosophy, foster new momentum with informatization, use new momentum to drive new development, and create new glories with new development [6]. In 2017, China wrote the digital economy into the government work report for the first time and incorporated it into the national development plan as a new driving force for improving the quality and efficiency of economic development [7]. In February 2023, the Central Committee of the Communist Party of China and the State Council issued the Overall Layout Plan for the Construction of a Digital China, which, from the overall and strategic height of the cause of the Party and the state, put forward the overall strategy for building a digital China in the new era and defined the guiding ideology, main objectives, key tasks and safeguard measures for that construction. Building a digital China is an important engine for advancing Chinese-style modernization in the digital age and a strong support for creating new national competitive advantages. The Chinese government's active promotion of digital transformation and data governance reflects the determination and boldness of the central leadership in building a digital China and

developing the data industry, all of which has strengthened the legal and policy foundation for data auditing. Such an efficient administrative environment and policy foundation are of great significance for guaranteeing the emergence and development of the data auditing industry and the timely application of data audit results.

2.2 Internal Weaknesses (W): Extensive Data Management and Incomplete Construction of Laws and Regulations

China is a country rich in data resources, and its data ecosystem has developed rapidly; yet this has been accompanied by the complexity and non-standardized nature of data management. Today, as China strides rapidly into the information age, the generation and collection of massive amounts of data has become the norm in daily life and business activities; however, the management and maintenance of these data face severe challenges. Specifically, China's data management environment is generally characterized by extensiveness and a lack of standards.

First, most organizations and enterprises lack clear data management strategies and norms. This means that the collection, storage, processing and sharing of data often lack unified standards and guidelines. Such a lack of standardization may lead to inconsistency of data between different departments and systems, which in turn affects the quality and credibility of data. Apart from technology enterprises with relatively advanced information management technology, traditional enterprises are generally rather extensive in data collection and management, and large numbers of data samples that should have become productive resources are wasted.

Second, the extensive management environment has brought data quality and data security problems to the fore. Owing to the lack of strict data management measures, data are often prone to such problems as error, redundancy, incompleteness and inaccuracy. This may not only lead to inaccurate decision-making but may also damage an organization's reputation and customer trust. At the same time, insecure storage and transmission of data may trigger data leakage and security vulnerabilities, further aggravating the risks.

What is worse, some enterprises deliberately leak users' information and data illegally in pursuit of profit, which already involves illegality, and many of them are leading enterprises in their industries. On July 2, 2021, the Cyberspace Administration of China launched a cybersecurity review of "DiDi". On the evening of July 4, the Cyberspace Administration of China issued a notice stating that, according to reports and after testing and verification, the "DiDi" app had serious problems of illegally and irregularly collecting and using personal information, and notified app stores to remove the "DiDi" app. On July 21, 2022, after nearly a year of investigation, the state announced the results of the investigation into DiDi and the penalties imposed. It was ascertained that DiDi was found to have illegally collected 11.9639 million pieces of screenshot information from users' mobile phone photo albums, and to have excessively collected 8.323 billion pieces of clipboard information and app list information. This is merely a relatively serious and typical case of personal information data leakage, and may be only the tip of the iceberg of the risks faced by personal data in China. A large amount of illegally leaked personal data flows to criminal organizations engaged in fraud and illegal usury, which has directly threatened the personal safety and property safety of the people who generate and provide the data.

Finally, data auditing, as a key activity for ensuring the reliability and compliance of data, faces serious challenges. Auditing requires reliable and accurate data as its basis in order to effectively supervise and verify an organization's data management practices. However, owing to the non-standardized nature and quality problems of data, auditors may find it difficult to obtain credible data, which affects the validity and reliability of audit results. China has a series of problems of an extensive and non-standardized management environment in data management, which not only threatens the quality and security of data itself but also poses challenges to the smooth conduct of data auditing. This is both a manifestation of the inherent contradiction between the short period of development and the limited accumulated experience in China's data management field and a rapidly and vigorously developing data industry, and an inevitable consequence of China's long-standing lack of attention to the protection of personal data, especially private information. Therefore, it is necessary to strengthen the standardization and supervision of data management

and to formulate clear data management strategies so as to ensure the effective use of data resources, the improvement of quality and the maintenance of security. This will help China better respond to the opportunities and challenges brought by the data age.

Although the Chinese government has already taken a series of measures to strengthen data governance and privacy protection, there are still many imperfections in the construction of legal rules for data auditing. This problem is particularly prominent in the current digital age, because data auditing requires not only rigorous methods and technologies but also a clear legal framework and rules to ensure the lawful acquisition, use and protection of data. Where legal rules are unclear or incomplete, data auditing may be seriously affected by legal risks and uncertainty.

First, it should be emphasized that the validity and legality of data auditing depend directly on a clear legal framework. This includes provisions on the lawful collection and processing of data, privacy protection regulations, and legal provisions on data access and sharing. At present, China's data law system has not yet been fully developed, and there exist some vague and undefined legal concepts, which brings challenges to data auditing. In China's current formal legal system, the only specialized laws concerning data protection are the Data Security Law of the People's Republic of China (hereinafter referred to as the "Data Security Law") and the Personal Information Protection Law of the People's Republic of China (hereinafter referred to as the "Personal Information Protection Law"), and there remains a problem of vague legal concepts between the two with respect to the most important issue of personal data protection. For example, the definition and scope of personal data still need to be clearly delimited. Article 38 of the Supplementary Provisions of the Data Security Law provides that where a state organ needs to collect or use data to perform its statutory duties, it shall do so within the scope of those duties and in accordance with the conditions and procedures prescribed by laws and administrative regulations, and that it shall keep confidential in accordance with law such data as personal privacy, personal information, trade secrets and confidential business information learned in the course of performing its duties, and shall not disclose them or illegally provide them to others. Article 53 provides that where data processing activities are carried out in statistical and archival work [...] after the means of processing personal information, does the data no longer fall within the category of "personal data"? Yet such data are truly the "energy" of the "data age". Should such data be included in the category of "personal data" and the rights of data subjects be protected? After the emergence of generative AI, the efficiency with which such personal data are used will grow exponentially, and where legal concepts remain vague, it will be very difficult to carry out the relevant supervision and protection work.

At the same time, the Data Security Law and the Personal Information Protection Law are conspicuously incomplete with respect to the protection of the property rights attributes of personal data. The Data Security Law focuses on protecting the value of security in data processing activities, especially the value of public security, and does not regulate the question of the ownership of property rights in data. The Personal Information Protection Law focuses on protecting personal information involving citizens' privacy and on constructing the relevant legal concepts of the right to personal information, mainly including the right of self-determination and control over personal information, as well as the right to know, the right to consult and copy, the right to correct and supplement, the right to delete, and the right to demand explanations. What the Personal Information Protection Law emphasizes is likewise the protection of the value of security, yet it has achieved nothing in building the discourse system of personal data rights. Personal data rights mainly include the right of access, the right to rectification, the right to erasure (the right to be forgotten), the right to restrict processing, the right to data portability, and the clarification of the property rights enjoyed by data subjects over data.

In an era in which "data is energy", the property value attribute of personal data is an important attribute of private rights, and China's legal supply in this regard is markedly insufficient. An insufficient supply of legal norms will severely restrict the scope of data audit activities and the development of the industry. Incomplete legal rules may lead to legal risks and uncertainty. In conducting a data audit, auditors need a clear understanding of the relevant

legal provisions in order to ensure that their conduct meets the requirements of the law. However, if legal rules are unclear or incomplete, auditors may face uncertainty about legal liability, because it is difficult for them to determine which conduct is lawful and which is unlawful. In the absence of clear rules, auditors may find it difficult to determine whether data processing meets the requirements of legality and compliance. This may lead to uncertainty in audit results and affect the credibility and validity of data auditing. Clear legal provisions can help ensure the legality and compliance of data auditing, reduce legal risks and uncertainty, and at the same time enhance the credibility and validity of data auditing. This is of great significance for safeguarding data security and the right to privacy and for promoting the sustainable development of the digital economy.

2.3 External Opportunities (O): The Rise of a New Round of Industrial Revolution and the Continued Growth of International Digital Trade

A new round of industrial revolution and the continued growth of international digital trade provide external opportunities for China in building data audit rules. China can make full use of these opportunities to strengthen the legal framework and standards for data auditing, promote the sustainable development of the digital economy, and improve data security and compliance so as to adapt to an increasingly digitalized economic environment. This will help further raise China's status and influence in the global digital economy.

The ongoing global wave of the new industrial revolution has brought about the worldwide rise and spread of such advanced technologies as AI, the Internet of Things, big data and cloud computing. The development of these technologies provides new opportunities for the development of the data auditing industry. As industries throughout the global industrial chain rely more and more on data-driven decision-making and operations, the importance of data auditing has increased markedly. China can ride the momentum of this new industrial revolution to promote the formulation of data audit regulations, ensure the lawful use and protection of data, and at the same time foster technological innovation and the sustainable development of the digital economy.

The rise of a new round of industrial revolution represents an era of transformation characterized by the convergence of advanced technologies. AI gives machines cognitive capability, the Internet of Things connects devices and collects large amounts of data, big-data analytics reveals valuable insights from enormous data sets, and cloud computing provides scalable and efficient data storage and processing capacity. The collective progress of these technologies has created an environment rich in data resources, thereby increasing the scope and complexity of data auditing. As industries rely more and more on data for critical decisions and operations, the need for effective data auditing has become more prominent. Data auditing serves as a key mechanism for assessing data quality, ensuring compliance and verifying the lawfulness of data use. In an age in which data breaches and privacy concerns are widespread, data auditing provides important protection against data abuse and misuse.

With its vibrant technological ecosystem and rapidly growing digital economy, China is fully positioned to harness the potential of the new industrial revolution and to advance the development of data auditing practice. By actively shaping data audit regulations and frameworks, China can build a solid foundation for the lawful use and protection of data. This will cultivate trust between enterprises and consumers and foster an environment conducive to technological innovation and the sustainable growth of the digital economy.

One of the most important impacts of the new round of industrial revolution on the pattern of international trade is the rise of digital trade. Digital trade is growing continuously worldwide, covering such fields as cross-border e-commerce, cloud services and digital content. As large generative AI models have become the core technology of the new round of industrial revolution, digital trade is of vital significance to the research, development and application of AI models in all countries. For China's large AI models to open up overseas markets, massive overseas data must serve as the basis for research and development. The emergence of large AI models fully highlights the significance of data as the "oil energy" of the new era. With the world's largest e-commerce market, China has the opportunity to play a key role in international digital trade. International digital trade enables goods, services and

information in digital form to be exchanged continuously across national borders. This trend covers various aspects, including the booming field of cross-border e-commerce, the provision of cloud services beyond geographical boundaries, and the global dissemination of digital content. Given that China is a global economic power thriving in terms of its digital ecosystem, it is fully positioned to make the most of its advantages and to take an active part in the thriving landscape of international digital trade.

Formulating and implementing robust data audit rules can enhance China's attractiveness to international partners and investors. By demonstrating a commitment to data transparency, privacy protection and regulatory compliance, China can attract foreign enterprises and stakeholders to participate in digital trade activities. This in turn may lead to an increase in foreign investment and cooperation, further raising China's influence in the field of global digital trade. The surge in international digital trade provides China with a compelling opportunity to elevate its key position in this vibrant field. Formulating data audit rules is crucial to ensuring the integrity and security of digital trade activities and can thereby give a strong boost to international cooperation and investment attraction and promote the sustainable development of the digital economy worldwide.

2.4 External Threats (T): U.S. Pursuit of Data Hegemony and the Rapid Development of European Data

Protection Rules

China likewise faces many external challenges in constructing data audit rules. The United States' pursuit of global data hegemony may obstruct China's efforts to establish its own data governance rules. As international data governance aligns more and more with U.S. interests and standards, China's ability to formulate its own rules and regulations may face growing challenges.

U.S. technology giants such as Google, Facebook and Amazon have accumulated enormous amounts of data from users worldwide, making them key players in the global data ecosystem. They possess tremendous influence over and control of global data. This situation poses challenges to China's construction of data trade and data audit rules, because it may not only restrict China's access to certain international data pools but, given these companies' extensive influence, also arouse China's concerns about data security and privacy.

The United States has made full use of its technological advantages to firmly establish global data dominance. By controlling data devices, terminals and transmission channels worldwide and defining data sovereignty as control over data, the United States has fundamentally challenged the sovereign boundaries of other countries and laid the foundation of its data hegemony. In addition, through control-centred standards, the United States has reshaped cyberspace as a globally shared resource and expanded its data sovereignty. This approach has met with opposition from other countries proceeding from their own independent rights and interests, which has become an obstacle to the construction of U.S. data hegemony. To overcome the opposition and obstacles of other countries based on sovereign interests, the United States has launched a new effort, attempting to export its liberal values in the name of creating a "free kingdom of data" so as to break through other countries' political defences. This strategy extends the American model of economic freedom into the data domain, effectively extending U.S. dominance from the traditional economic domain to the data domain [11].

As mentioned earlier, the United States has undisputed advantages in global data terminals and transmission channels, and therefore possesses obvious advantages in data collection and utilization. If foreign governments restrict cross-border data flows through administrative means, the United States' ability to access global data will be weakened. Therefore, by using such instruments as treaties to impose the free flow of cross-border data, the United States can continue to safeguard its interests and strengthen its data hegemony. The construction of U.S. data hegemony has been seamlessly integrated into the existing hegemonic system and, relying on its formidable strength, has advanced along the "institution-value" approach. However, when confronted with attempts by foreign countries to obtain U.S. data, the United States uses legal instruments in such fields as trade, technology and finance to intervene in the countries and enterprises concerned. Such intervention is intended to prevent other countries from infringing

on U.S. data and from threatening its data hegemony, which highlights the existence of double standards. All of this constitutes an important challenging factor for China in building a data audit rule system with international influence.

Besides the United States, the rapid development of European data protection rules is another external challenge. Through the implementation of the General Data Protection Regulation (hereinafter referred to as “GDPR”), Europe has become a leader in global data protection rules. The GDPR sets strict standards for data privacy, security and consent, and has influenced how organizations worldwide handle the data of European citizens. The rapid development and strict enforcement of these rules pose a challenge to China’s data audit rules. In addition, the GDPR has already initially acquired the status of a legal basis and model norm for international data auditing: Article 3 gives the Regulation itself extraterritorial jurisdictional effect, and countries including the United Kingdom, India, Brazil, South Africa and Australia have all drawn inspiration from the extraterritorial legal framework of the GDPR. Following the example of the GDPR, they have attempted to extend the extraterritorial scope of their own data legislation. This clearly shows that extending the jurisdiction of national data legislation beyond national borders has become a common practice in many countries, and the phenomenon of extraterritorial jurisdiction in the field of data legislation has thus evolved into a global trend [12].

3 Strategies for Building China’s Data Audit Rules

Based on the SWOT model analysis above, this paper proposes four strategies for China in building data audit rules—the growth strategy (S-O), the sustaining strategy (S-T), the turnaround strategy (W-O) and the defensive strategy (W-T)—in the hope of providing strategic support for China to build a data audit rule system with international influence during the window period of the rise of the large generative AI model industry.

3.1 Growth Strategy (S-O)

Viewed from a global perspective, a new round of industrial revolution means that China’s external opportunities are not limited to seizing the wave of the industrial revolution to promote internal development; on the basis of China’s existing advantages, greater global opportunities have already emerged. China already has the strength to attempt to redefine its international identity in the new round of industrial revolution. Specifically, China’s opportunities include but are not limited to shifting from a beneficiary of technological application to a leader in technological research, development and innovation, and from a follower of international technical standards and international technical management rules to a maker of them. Data audit services and AI technology are both emerging new fields. In this important window period, China should focus as early as possible on building and improving the industrial policies and the industrial legal basis for data auditing and, while improving domestic legal norms, should also devote itself to developing data audit norms that can serve global data flows and trade, so as to occupy a favorable position in the development of data audit norms at the level of international law.

3.2 Sustaining Strategy (S-T)

China should make full use of its own advantages in data resources and the digital industry to address such challenges as U.S. data hegemonism and Europe’s strict data protection rules, fully carry out trade negotiations with the West, and coordinate as far as possible the conflicting interests of the parties in the field of data trade, taking the bilateral and multilateral trade agreements eventually reached as the legal basis for China’s international data auditing so as to promote the stable development of China’s international digital trade; the relevant achievements in rule-making at the level of international law can all become sources of rules for improving domestic law.

3.3 Turnaround Strategy (W-O)

China should shift the emphasis in building the data audit rule system from a one-sided protection of the value of security to an all-dimensional protection of data subjects. It should provide legislative clarification of important concepts that are still in a vague state in the current legal system, such as “personal data”, and should strive to build a system for protecting the rights of data subjects, especially rights of a private nature. By improving legal rules,

it should promote the penetration of data audit activities into every link of data application, give full play to the market-oriented, routine and professional characteristics of data audit activities, better, more comprehensively and more deeply protect the property rights that data subjects should enjoy, foster an awareness of data rights throughout society, and encourage Chinese producers to develop the habit of refined data management.

The formulation of data audit rules based on the protection of property rights can enhance China's competitiveness in the field of digital trade, attract more international partners and investors, and promote international cooperation and development in the digital economy. China's status as a major participant in the field of international digital trade highlights the importance of establishing comprehensive data audit rules. Such rules are the cornerstone for ensuring data integrity, security and conformity with international trade norms. Effective data auditing practice not only safeguards the interests of consumers and enterprises but also helps raise the overall credibility and reliability of China's digital trade activities on the global stage.

3.4 Defensive Strategy (W-T)

China should respond at the legislative level to challenges worldwide as soon as possible. With regard to the hegemonic conduct of the United States, China should persist in building data audit and management rules based on sovereign interests, resolutely respond by legal means to the erosion of data sovereignty by the United States under double standards, and, in international digital trade activities, adhere to the basic principle that the application of audit rules must be based on sovereign interests. Coordinating China's regulatory framework with European standards is a highly complex task that requires careful consideration of the principles of privacy and security. Failure to meet these evolving international data protection standards may impede China's ability to conduct cross-border data exchange and cooperation with European partners. In response to the system of extraterritorial data jurisdiction established by Europe, Article 3 of China's Personal Information Protection Law draws on the legislative approach of the GDPR and creates the extraterritorial jurisdiction system of China's Personal Information Protection Law. Although China's legislative response cannot be said to be untimely, the relevant norms do not yet have the capacity to become a legal model in the field of international data auditing, and their scope of jurisdiction is too narrow, so the relevant legislative work still has a long way to go.

4 Conclusion

As a new type of risk brought about by the development of large AI models, data risk faces a predicament of insufficient legal and institutional supply in both its governance and its response. Data auditing, as a market-oriented, routine and professional management instrument, has become a feasible option for responding to such risks. China should further improve and build the legal rule system for data auditing and formulate relevant policies to drive and promote the development of the data auditing industry, so as to achieve its own industrial upgrading and leapfrog development in the new wave of industrial revolution and to seek a more favorable position amid changes in the international landscape.

References

- [1] Y. M. Liu, L. Ma, Y. Sun et al., "The challenges and opportunities of ChatGPT for public governance (a written discussion)," *Dongfang Forum*, no. 3, pp. 1-24, 165, 2023.
- [2] Y. W. Chen, "Beyond ChatGPT: Opportunities, risks and challenges of generative AI," *Journal of Shandong University (Philosophy and Social Sciences)*, no. 3, pp. 127-143, 2023.
- [3] X. D. Dou, "Risk and control: On the protection of personal information in the application of generative artificial intelligence," *Tribune of Political Science and Law*, no. 4, pp. 59-68, 2023.
- [4] X. Zhang, "Data risks of generative artificial intelligence and the path of governance," *Science of Law (Journal of Northwest University of Political Science and Law)*, no. 5, pp. 42-54, 2023.

- [5] N. Shen, "A SWOT model study on the digital communication strategy of intangible cultural heritage," *Journal of Tongji University* (Social Science Edition), no. 1, pp. 58-67, 2023.
- [6] X. Gao, "Establishing a new form of government governance adapted to the digital age," *Exploration and Free Views*, no. 4, pp. 141-146, 179-180, 2021.
- [7] Z. H. Li, "Research on the impact of digital transformation on enhancing enterprise value and its transmission path," *On Economic Problems*, no. 11, pp. 25-32, 2023.
- [8] L. Tong, "Lessons from the 'DiDi' incident: Enterprise data security compliance must hold the red line," *China Information Security*, no. 7, pp. 84-85, 2021.
- [9] L. J. Liu, "Differentiation and analysis of personal information and personal data," *Probe*, no. 5, pp. 151-159, 2022.
- [10] H. D. Wang, "Research on the approach to the propertization of personal data," *Administrative Law Review*, no. 6, pp. 90-100, 2021.
- [11] S. Feng, "The construction of U.S. data hegemony and the legal response," *Pacific Journal*, no. 8, pp. 45-58, 2023.
- [12] X. Wang and W. Shi, "The globalization of extraterritorial jurisdiction in data legislation and China's response," *Intellectual Property*, no. 4, pp. 54-75, 2022.