Asia Pacific Science Press

Asia Pacific Economic and Management Review
ISSN(O):3005-9275
ISSN(P):3005-9267
Vol. 3 No.2 (2026)

Research on the Construction of Digital Standard Contracts for Artificial Intelligence Data Training

Qiuhui Ren[1]*, Xiaohua Fu[2]

1. School of Marxism, Henan Forestry Vocational College, Luoyang, Henan, 471002, China
2. School of Cultural Tourism, Chengdu Polytechnic, Chengdu, Sichuan, 610041, China

Corresponding author: Qiuhui Ren, qiuhuiren0907@126.com

Abstract: Based on the analysis and comparative study of domestic and foreign literature, this paper focuses on the legal regulatory issues arising from the process of artificial intelligence (AI) data training. It aims to address the inherent contradictions between machine learning models and the technical characteristics of traditional legal norms, the lack of algorithmic transparency, and the identification of data ownership by constructing digital standard contracts. Corresponding development suggestions are put forward from three aspects: establishing a dynamic update mechanism for contracts, standardizing transnational data training cooperation, and deepening the legal effect of contract execution results.

Keywords: Artificial Intelligence; Data Training Behavior; Digital Standard Contract
Published: Apr 12, 2026
DOI: https://doi.org/10.62177/apemr.v3i2.1246

1. Introduction

1.1 Research Background

With the rapid development and wide application of AI technology, data training, as the core link of AI systems, involves the collection, processing and utilization of massive data, triggering a host of new legal issues in the legal field. Risks such as data privacy leakage, discriminatory results caused by algorithmic bias, and disputes over intellectual property ownership have become increasingly prominent, exposing the practical dilemma of the lack of effective legal regulation on AI data training behaviors[1,2]. In addition, from intelligent judicial systems to automated administrative decision-making, AI has penetrated into all aspects of legal practice. While improving efficiency, this technological application has also brought about many legal challenges. There is no unified global standard for data training, and the legal conflicts among different jurisdictions and the limitations of industry self-regulatory norms have become increasingly prominent[3]. These changes brought about by AI are triggering a profound transformation in the legal field. Against this background, it is particularly urgent to construct binding and universal digital standard contracts, which is not only a key measure to balance technological innovation and rights protection, but also an important path to realize the modernization of AI governance.

At present, most academic research on AI regulation focuses on the macro-policy level, with insufficient attention paid to specific operational standards. In the field of data training, existing norms are fragmented[4]. Industry self-regulatory standards lack enforce-ability, and international standards are difficult to adapt to the local legal environment. This lack of norms leads to high compliance costs for enterprises and insufficient law enforcement basis for regulators, ultimately affecting the healthy

development of AI technology. Therefore, constructing a systematic and enforceable digital standard contract is not only of theoretical innovation value, but also can provide practical solutions for practice. Research on the governance of the artificial intelligence industry stands at a critical transformative juncture. It boasts a wealth of massive application scenarios and clear policy support, yet the current research capacity has not fully translated these advantages into academic and industrial strengths[5]. Going forward, it is imperative to closely follow the four major trends of scenario-based deep cultivation, standard formulation, paradigm transformation, and ecological collaboration, striving to build a nationally distinctive and influential research hub for applied artificial intelligence governance.

1.2 Construction Principles

The Balance Between Compliance and Development Drawing on the open data-set principles proposed by Mozilla and the requirements of the national AI standard system, the agreement is developed in accordance with five core principles:

1. Legality and Compliance First

Convert requirements such as “informed consent” and “minimum necessary” stipulated in the “Personal Information Protection Law” into mandatory contractual clauses. For instance, personal data used for training must be accompanied by proof of data subject authorization, and sensitive data must complete a Data Protection Impact Assessment (DPIA).

2. Equivalence of Rights and Obligations

Clarify the boundaries of rights and responsibilities among data providers, trainers, and users. This safeguards the data subject’s right to be forgotten”， a built-in data deletion trigger mechanism in the agreement while protecting the algorithmic intellectual property rights of trainers.

3. Technology-Neutral Adaptation

Be compatible with various privacy-enhancing technologies including federated learning and differential privacy. With reference to GB/T 46284-2025 Artificial Intelligence—Technical Specifications for Federated Learning, no specific technical solution is mandated; only security performance indicators are defined.

4. Openness, Collaboration and Sharing

Establish standardized data metadata specifications, support cross-institutional agreement mutual recognition, facilitate the construction of a competitive AI ecosystem, and lower the barrier for small and medium-sized enterprises to participate.

5. Dynamic and Controllable Risks

Introduce a dynamic update mechanism for contractual clauses, enabling real-time optimization of governance rules in response to policy adjustments (e.g., changes in cross-border data transfer rules) and technological developments (e.g., emergence of new attack methods).

2. Concepts and Research Status

2.1 Concepts

As an interdisciplinary subject integrating computer science, mathematics, cognitive science, linguistics and other fields, AI has made remarkable progress in theoretical research and practical application in recent years. Data training, as a key link in the development of AI models, is mainly characterized by the process of data collection, cleaning, annotation and model training, which directly affects the performance and application effect of the model. High-quality data training behavior can significantly improve the performance and generalization ability of the model[6]. However, data training behavior also faces some challenges, such as data privacy protection and model inter-pretability.

A digital standard contract refers to a contract created, executed, managed and stored using digital technology, containing pre-set clauses, with high consensus, oper-ability and exemplary characteristics, realizing machine readability, executability and automation, and with standardized clauses and data structures. Technically speaking, a digital standard contract integrates part of the mechanism of “technology-driven measures” into the clauses of “legal-driven measures” by using Boolean operators, formulating a “digital standard contract” with high flexibility to meet the needs of multiple scenarios. It can solve complex problems in the process of AI data training, such as fair use, licensing authorization, text and data mining exceptions, rights protection, data sharing, and tort liability division. It can make up for the lack of rules and reconcile the different demands of all stakeholders involved in AI data training to the greatest extent.

AI, data training behavior and digital standard contract have a close synergistic development relationship. The progress of AI technology relies on high-quality data training behavior, and the standardization of data training behavior requires the support of digital standard contracts. Therefore, the formulation and implementation of digital standard contracts can provide a unified specification for the application of AI technology, thus promoting the wide application of AI technology[7]. On the contrary, the development of AI technology also provides new tools and methods for the formulation of digital standard contracts, such as automatically generating standard clauses through generative AI tools.

The core value of digital standard contract research is reflected in three dimensions: Firstly, it realizes the organic unity of technical ethics and legal compliance by establishing a full-process specification for data collection, processing and use[8]. Secondly, the pioneering "dynamic compliance" mechanism of the contract enables technical standards to be updated synchronously with legal revisions. In the research of the financial field, clarifying the rights and responsibilities of data providers, algorithm developers and regulators can greatly reduce the blind spots of traditional financial supervision. Finally, the contract innovatively introduces a "technology-legal" dual-track verification system[9]. Its copyright filtering module not only achieves a higher accuracy of infringement identification, but also forms an evidence fixation standard in line with the judicial interpretation of the Copyright Law.

2.2 Research Status

Current research suffers from the following shortcomings:

1. Disconnection between disciplines.

Scholars in law and ethics lack in-depth understanding of the underlying technical principles of artificial intelligence, Transformer architecture for large models, reinforcement learning[10]. As a result, their discussions remain superficial and principlebased, fairness should be ensured, failing to propose operable governance solutions that can be embedded into the technical foundation, how to achieve fairness through technical means.

2. Limited practical impact of research outputs. Many findings remain at the level of academic papers and reports, without effective translation into concrete policy provisions, industry standards, or practical guidelines for enterprises.

3. Insufficient industryacademia interaction.

Except for a few leading enterprises such as Yutong, universities and research institutes have limited engagement with most small and mediumsized AI enterprises in the province. Theoretical research has not adequately responded to the practical difficulties encountered by enterprises in their development.

Based on existing research capacity, the regulated development of artificial intelligence industry will exhibit five major trends in the future:

1. From principle discussion to scenariobased deepening: refinement and verticalization. Macrolevel, allencompassing narratives will be gradually abandoned. Research will deeply dive into specific industry scenarios to produce highly operable governance schemes.

2. From theoretical output to standardsetting: pursuing discourse power and practicality[11]. Efforts will be actively made to translate theoretical achievements into local standards, industry standards, and even national standards.

3. From humanitiesdriven to technologydriven: paradigm shift in research. More studies will leverage privacy computing, block-chain, federated learning, and other technologies to promote the regulated development of the industry—using technical solutions to address governance challenges. The focus will shift from ex post regulation to ex ante prevention, advancing the implementation of "ethics by design" and "compliance by design" in AI product development workflows.

4. From singleparty dominance to governmentindustryuniversityresearch collaboration: ecological integration and closedloop governance. Industry will provide scenarios and data; enterprises and universities will jointly build laboratories with realworld data and application scenarios for empirical analysis and experimental validation[12]. Universities will offer theoretical support and solutions while cultivating and delivering interdisciplinary talents for enterprises, forming a closedloop system.

5. From riskfocused to developmentbalanced: shift in research orientation. Research objectives will no longer be limited to mere "control" and "risk avoidance," but to maximizing the innovative development of the AI industry under effective

governance.

2.3 Research Approach

The construction of digital standard contracts needs to break through the traditional legislative thinking. At the methodological level, an innovative path of "legal encoding" should be adopted to transform abstract legal principles into enforceable technical parameters[13]. For example, privacy protection requirements can be quantified into specific indicators of data desensitization, and algorithmic fairness can be transformed into threshold standards for model deviation. In addition, in terms of content design, the contract should establish a full-chain normative system: clarify the authorization method and scope restrictions in the data collection stage; stipulate algorithm audit and record retention requirements in the training stage; and establish a liability tracing and relief mechanism in the application stage. This three-dimensional normative framework can effectively cover the entire life cycle of data training.

The design of research methods should focus on the organic combination of empirical and comparative research. By analyzing controversial cases in typical application scenarios such as medical diagnosis and credit evaluation, the legal risk points of data training can be accurately identified. Comparative law research helps to draw on the legislative experience of the EU AI Act, the US Algorithmic Accountability Act and other legislations, and carry out localized adjustment combined with the characteristics of China's legal system[14]. The technical verification link needs to design a multi-level test plan, including single-point technical testing, system integration testing and stress testing, to comprehensively evaluate the technical feasibility of the contract. This multi-dimensional research method can ensure the close connection between theoretical construction and practical needs[15].

The implementation path needs to adopt a progressive strategy. In the short term, industry alliances can be used to promote the formation of best practices, such as establishing an authentication mechanism for the source of training data. In the medium term, a third-party certification system can be developed to grant compliance certification to data training behaviors that meet the standards. In the long run, efforts should be made to promote the upgrading of core standards to mandatory norms and coordinate national legislation through international organizations. In particular, a dynamic update mechanism should be established to enable standard contracts to adapt to the rapid iteration of technology. The standard-setting model of the Internet Engineering Task Force can be used for reference to establish an open and transparent contract revision procedure to ensure that all stakeholders can participate in the improvement of rules[16].

Where laws remain unclear, "digital standard contracts" can serve directly as an effective tool for resolving complex issues involved in artificial intelligence data training. They meet the urgent need for effective regulation of data training practices by individuals, market regulators, AI R&D and application enterprises (such as Deep Seek), data aggregation platforms, and other stakeholders. The promotion and translation path of digital standard contracts is well-defined and can deliver tangible results. The process of researching and designing the principles, clauses, and other elements of standard contracts also expands the research approaches to core AI-related issues—including subject status, rights and interests, and tort determination—from a brand-new, bottom-up, and highly practice-oriented perspective. This is of great value in promoting the regulated development of the artificial intelligence industry.

3. Problems

The construction of digital standard contracts for AI data training behaviors faces multiple dilemmas. From the perspective of technical characteristics, there is an inherent contradiction between the iterative characteristics of machine learning models and the stability of traditional legal norms. Taking the EU General Data Protection Regulation as an example, its stipulated purpose limitation principle requires that data processing must be consistent with the purpose of initial collection, while machine learning models often derive application scenarios beyond the original scope in the continuous training process[17]. This tension is particularly prominent in highly sensitive fields such as medical care and finance. Medical institutions often face compliance risks when using patient data for algorithm optimization. In addition, transnational technology companies carrying out data training activities globally need to cope with the differences in data cross-border flow rules among different jurisdictions, which further increases the complexity of legal application.

The lack of algorithmic transparency is a technical bottleneck hindering effective legal regulation[18,19]. The "black box"

characteristic of deep neural networks makes the model decision-making process difficult to trace. When an algorithm produces discriminatory results, victims often find it difficult to provide evidence. In a 2021 case of algorithmic discrimination on a US recruitment platform, the plaintiff lost the case because they could not obtain the specific parameters of algorithm training. This case highlights the inadequacy of the existing legal relief mechanism[20]. More notably, with the rise of multi-modal large models, the complexity of training data sources has increased exponentially, making algorithm audits face unprecedented technical challenges. The algorithm impact assessment system under the current legal framework is often difficult to cope with such a large-scale and complex training system.

The identification of data ownership urgently needs legal clarification. Under the traditional copyright law system, the acquisition and use of training data involve multiple rights relationships[21]. For example, public data crawled by web crawlers may contain copyright-protected content, and the applicable boundary of the current fair use system in machine learning scenarios is not clear. A 2023 infringement lawsuit against a well-known AI painting tool embodies this contradiction, where the plaintiff claimed that their artistic works were included in the training data set without permission[22]. At the same time, disputes over the ownership of AI-generated content are increasing, and there are obvious differences in judicial practices in various countries. Some countries recognize the copyright of AI-generated content, while others adhere to human author-centered doctrine. This legal uncertainty seriously restricts the rational flow and utilization of data resources.

The existing normative system has structural defects. In terms of normative form, industry self-regulatory standards lack enforceability, while international treaties with legal binding force are difficult to coordinate the interests of various countries. Although the AI ethics guidelines issued by the International Organization for Standardization (ISO) provide a technical framework, they do not involve specific division of legal liability[23,24]. In terms of normative content, most existing rules focus on the data collection link, and the specification of the training process itself is relatively weak. Taking face recognition technology as an example, although many countries have legislated restrictions on the collection of biometric data, there is a lack of detailed provisions on data processing behaviors in model training. This regulatory gap leads to the lack of clear guidelines for enterprise compliance and also makes regulatory law enforcement face the dilemma of inconsistent standards.

4. Measures

Establish a dynamic update mechanism for contracts It is recommended to set up a permanent revision committee composed of experts in law, computer science, ethics and other disciplines, and formulate a quarterly evaluation cycle to ensure that contract clauses can respond to technological iteration and legal revisions in a timely manner[25]. At the same time, a rapid transformation mechanism for contract clauses should be established to convert newly promulgated laws and regulations into enforceable technical parameters in the shortest time. This dynamic update mechanism needs to be supported by the development of automated monitoring tools to track the changes of relevant legislation in major jurisdictions around the world in real time.

Standardize transnational data training cooperation At present, there are significant differences in national laws and regulations on data sovereignty, privacy protection and other aspects, which brings compliance challenges to cross-border AI research and development[26]. It is recommended to develop contract variant modules with regional adaptability on the basis of the existing contract framework. For example, data sovereignty protection clauses can be strengthened for the EU region, and digital copyright protection mechanisms can be focused on for Southeast Asian countries. At the same time, it is necessary to explore the establishment of a transnational certification mechanism and realize the mutual recognition of standards among different jurisdictions through bilateral or multilateral contracts[27].

Deepen the research on the legal effect of contract execution results. Focus on solving the issue of the admissibility of materials such as algorithm audit reports and compliance certifications in judicial proceedings[28]. It is recommended that the Supreme People's Court issue relevant judicial interpretations to clarify the criteria for determining the probative force of digital standard contracts in litigation. At the same time, a unified algorithm filing and registration system should be established to provide a basic basis for subsequent liability determination. The improvement of these systems will further enhance the legal authority and enforceability of contracts.

The innovation points of this research mainly lie in the methodological level: First, it proposes a transformation path

of "legal requirements technologization", decomposing abstract legal principles into quantifiable algorithm parameters, such as concretizing the "minimum necessity principle" in the Personal Information Protection Law into a threshold of access frequency for data fields. Second, it has developed a "compliance embedding" technical architecture, enabling legal requirements to directly influence the neural network training process, achieving a balance between algorithm bias correction and model performance maintenance in tests. Third, it has established an interdisciplinary collaboration mechanism, combining compliance datasets labeled by legal experts with verification tools developed by technical personnel, solving the problem of legal lag behind technological development in traditional governance. These innovations ensure that the protocol maintains the authority of legal norms while also possessing the feasibility of technical solutions.

5. Conclusion

The research points out that digital standard contracts can provide an institutional solution that balances technological innovation and rights protection for AI data training behaviors through the innovative path of "technicalization of legal elements". Its dynamic update mechanism and cross-border adaptability design are more exemplary for building a global AI governance system.

In the long run, digital standard contracts are expected to develop into a new type of legal infrastructure in the digital era. With the deepening of research, the application scope of digital standard contracts can be expanded from the data training link to the full life cycle management of AI[29]. In the future, it is possible to explore the integration of the contract framework with emerging technologies such as block-chain and privacy computing to build a more intelligent and transparent compliance governance system. This will not only help regulate the development of AI, but also provide an important reference for the construction of digital rule of law.

Funding

2026 Annual Key Research and Promotion Special Project (Soft Science) of Henan Province {Research on the Digital Standard Contracts for Artificial Intelligence Data Training) 262400410627

Conflict of Interests

The authors declare that there is no conflict of interest regarding the publication of this paper.

Reference

[1] Russell, S. J., & Norvig, P. (2021). Artificial intelligence: A modern approach (4th ed.; J. Yin, Trans.). Tsinghua University Press.

[2] Guo, D., & Zhang, Y. (2024). Infringement risks of generative AI training data and legal responses. Journal of Xiangtan University (Philosophy and Social Sciences), 48(5), 78–86.

[3] Jiang, H., & Tan, X. (2026). Beyond techno-nationalism: Governance dilemmas and regulatory approaches to international AI security. Contemporary China and World, (1), 65–75.

[4] Wu, Y., Chen, Y., Yang, H., et al. (2024). Fragmentation of global digital governance under algorithmic institutional competition and its transcendence path. Modern Distance Education, 1–15.

[5] Li, M. (2024). A report on the translation practice of international contract law: AI, data, cybersecurity and the legal landscape from the perspective of teleology [Master's thesis]. Guangdong University of Foreign Studies.

[6] Gao, X. (2025). Digital empowerment of dual contract risk control to improve the operational resilience of manufacturing enterprises. Information Construction, (6), 56–57.

[7] He, C., & Ma, G. (2025). Analysis of computer communication technology and electronic information technology in the field of AI. China New Communications, 27(20), 16–18.

[8] Liu, X. (2024). "Non-work use" in generative AI data training and its legitimacy justification. Legal Forum, 39(3), 67–78.

[9] Zuo, S. (2026). Validity identification and dispute resolution path of electronic contracts in the context of digital economy. Legal Vision, (3), 52–54.

[10] Crawford, K. (2021). The atlas of AI. Yale University Press.

[11] Hu, X., & Zhou, Y. (2025). A review of research on economics and management disciplines based on generative AI. Chinese Journal of Management Science, 33(1), 76–97.

[12] Cheng, X. (2026). On the inapplicability of network infringement rules to infringements of generative AI services. Comparative Law Review, (1), 125–137.

[13] Cheng, Y., Chen, G., Chen, H., et al. (2022). Exploration on key technical paths of standard digitization based on AI. Information Technology and Standardization, (10), 60–67.

[14] European Union. (2024). AI Act (EU AI Act).

[15] China Academy of Information and Communications Technology. (2023). White paper on AI data training security.

[16] Liu, S., Zhou, L., Yang, J., et al. (2022). Evolution of AI industry technology standard cooperation network and subject identification: Based on social network analysis and TOPSIS entropy weight method. Science and Technology Management Research, 42(6), 14–152.

[17] Ma, S., Yi, Z., Pan, G., et al. (2026). Host country data privacy protection policies and China's digital service trade exports: Evidence from the EU General Data Protection Regulation. Finance and Economics, 42(1), 39–49.

[18] European Commission. (2023). Guidelines on algorithm impact assessment.

[19] Gao, Q. (2023). Legal realization path of algorithmic transparency. Journal of Political Science and Law, (4), 112–125.

[20] Wu, Z. (2025). Justification of fair use path for AI data training: A comment on the first fair use cases in China and the United States. Journal of Shandong University of Science and Technology (Social Sciences), 27(6), 53–61.

[21] Gao, Y. (2024). Regulation of copyright infringement by AI training data. China Publishing Journal, (15), 12–18.

[22] Zhou, H. (2026). Judicial protection of AI models: From the perspective of unfair competition dispute case of “Transformation Comic Special Effect”. Journal of Law Application, (3), 71–86.

[23] Shen, K. (2024). On the implementation mechanism of soft law: Taking AI ethical norms as an example. Finance and Economics Law Review, (6), 108–127.

[24] IEEE. (2021). Standard for ethical aligned design (IEEE 7000-2021).

[25] Liu, Y. (2026). How to achieve new breakthroughs in China's AI legislation? Practical solutions based on security, rights and governance. Law and Social Development, 32(2), 187–207.

[26] Song, Y. (2025). Legal risks and regulatory paths of generative AI data cross-border flow. Cybersecurity and Data Governance, 1–8.

[27] Cai, X. (2023). A comparative study on AI governance between China, the United States and Europe. Legal Forum, (3), 88–95.

[28] Shen, F. (2025). Risks to judicial justice arising from the development of AI justice and its prevention and governance. Legal Research, 1–15.

[29] Hu, B., Wu, J., & Zhang, S. (2024). Research on AI knowledge creation: Overall framework and future prospects. Journal of Hangzhou Dianzi University (Social Sciences), 20(5), 26–39.