Proceeding of ICGPSH 2025 Symposium: International Relations and Global Governance
DOI: 10.54254/2753-7048/2025.BO23202

Fair Use of Training Data in Generative Artificial Intelligence

Weiyi Xia

Guanghua Law School, Zhejiang University, Hangzhou, China
2572049411@qq.com

Abstract. Generative AI data training poses novel challenges to the doctrine of fair use in copyright law. While the "three-step test" can preliminarily justify the legitimacy of AI data training as fair use, it still faces deep-seated contradictions such as the disconnect between current legal frameworks and technological advancements, an overemphasis on "rights protection",and imbalances in copyright interests. Therefore, it is imperative to reform and innovate the fair use system. First and foremost, a copyright system centered on "fair use" should be established to break free from the constraints of "author centrism." In addition, a specific clause for "AI learning and creation" should be introduced to explicitly recognize data training as fair use. Furthermore, by integrating the four-factor test from U.S. copyright law, a general standard for assessing fair use should be established to balance copyright protection with technological development, thereby advancing the adaptation of copyright law to the demands of the digital era.

Keywords: Generative artificial intelligence, Data training, Fair use, Three-step test

1. Introduction

In recent years, artificial intelligence (AI) has rapidly transformed from a scientific fantasy into a powerful technological force with profound impacts in the real world. Whether it's AI systems in healthcare assisting doctors in diagnosing diseases through the analysis of vast amounts of medical imaging data, or AI algorithms achieving more precise operations and quality control in industrial production, all these demonstrate its immense potential and application value hidden beneath modern society. From a technical perspective, machine learning, as the core of AI, enables neural networks to process massive amounts of data and uncover valuable patterns and rules. A large amount of data is used to train these models, ranging from simple text data to complex image, audio, and video data, making data the key fuel for AI development. In this process, data training, as the foundation of AI technology development, has increasingly highlighted issues related to its legality and compliance.

In the field of copyright law, the issue of fair use in AI data training has become a focal point for both academia and practice. During the process of AI data training, the reproduction, analysis, and use of large amounts of works are inevitable. This may include texts, images, audio, and video works protected by copyright law. Due to the unique nature of artificial intelligence technology, it requires a substantial amount of diverse data, often necessitating deep-level analysis and processing, which poses significant challenges to traditional copyright protection models. On one hand, copyright law aims to protect the intellectual achievements of creators and prevent their legitimate rights from being infringed; on the other hand, the development of artificial intelligence relies

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-7048/2025.BO23202

heavily on vast amounts of data, and overly stringent copyright protections could also hinder technological progress and application.

At the international level, countries have adopted diverse legislative and policy measures regarding copyright issues in AI data training. The EU has provided exceptions for text and data mining through the Directive on Copyright in the Digital Single Market [1], Japan has added provisions on "information analysis" to its copyright law, and the United States has interpreted fair use through judicial precedents [2]. However, although China's revised Copyright Law in 2020 included "other circumstances prescribed by laws and administrative regulations" in the fair use clause, moving from a "closed" to a "semi-closed" restriction and exception model, it does not explicitly address specific situations involving AI data training. In May 2023, the Cyberspace Administration of China issued the Interim Measures for the Administration of Generative Artificial Intelligence Services, which initially form the basis for regulating generative AI training datasets under Articles 6, 7, and 8, where some provisions remain relatively vague. For example, "Article 7: Providers of generative artificial intelligence services (hereinafter referred to as providers) shall legally conduct pre-training, optimization training, and other data processing activities for training purposes, in compliance with the following rules: (2) Involving intellectual property rights, they must not infringe upon others' legally enjoyed intellectual property rights;" but it does not specify clear standards or methods for determining infringement. AI training data involves large amounts of text, images, audio, etc., and how to accurately determine whether there is an infringement during training and application, such as whether secondary creation or adaptation of data constitutes infringement, lacks practical guidelines [3]. This makes it difficult for AI enterprises to face great legal risks and uncertainties when conducting data training in practice.

Currently, the academic community generally pays less attention to AI data training. The copyright issues involved mainly focus on: whether the extensive use of data materials, including copyrighted works, in the learning and creation processes of AI constitutes fair use?[4] If it does, what institutional challenges arise? Based on this, this article aims to explore the issue of fair use of copyrights in AI data training, analyze the legality of AI data training under the current copyright law framework, and propose corresponding solutions. This is intended to protect the legitimate rights and interests of copyright holders and promote the healthy development and application of artificial intelligence technology.

2. The technical principles of generative AI data training

Machine learning has not formed a unified definition in academia. Arthur Samuel believes that machine learning is a research field that gives computers the ability to learn without obvious programming [5].

From an outcome-oriented perspective, machine learning can be categorized into "expressive" and "non-expressive" types, depending on whether the machine will eventually output expressive content. From this classification, The input and learning phases are "non-expressive," while the output phase is "expressive."

The essence of AI training is machine learning, which relies on large datasets to improve model performance and better respond to user commands. While the learning process is important, machine learning heavily relies on the quality and quantity of training data [6]. AI data training methods include supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning. Supervised learning improves the model by using known data and its outputs, continuously adjusting parameters to enhance performance. Unsupervised learning trains with data that does not have labels or outputs which are defined in advance. Semi-supervised learning combines features of supervised and unsupervised learning, using labeled data with known categories and unlabeled data without known categories for training, thus completing recognition

DOI: 10.54254/2753-7048/2025.BO23202

tasks. Reinforcement learning is about AI machines taking a series of actions in an environment and selecting best behavior strategies based on environmental feedback signals.

Except for reinforcement learning, the training process involves data collection, preprocessing, labeling, input, model training, and output [7]. Data collection involves obtaining large amounts of raw data for model training, primarily through public datasets and self-built datasets. Public datasets can be accessed under license agreements, reducing training costs; self-built datasets are constructed via online platforms or by creating them independently. After collection, the data needs to undergo preprocessing to remove harmful elements from the training data, ensuring consistency. This includes operations such as noise reduction, augmentation, and contrast enhancement for image data, improving data quality to better meet the requirements of model training.

Data annotation primarily involves manually labeling training data, and whether the annotation process is required depends on the learning model of artificial intelligence. In (semi-)supervised learning models, data must be annotated for AI models to learn and recognize patterns in the labeled data, whereas unsupervised learning does not require any annotations. During the data input phase, AI developers store processed data on servers for the model to read, which then adjusts parameters through training. Model training, as a cyclic process of inputting data, aims to find the optimal model state where the loss function is minimized by continuously adjusting model parameters, ultimately producing the desired output. Data output can vary depending on whether the AI is decision-making or generative. For example, decision-making AI like dog and cat recognition models typically produce image classification results. Generative AI, such as ChatGPT, may generate original content based on input data and trained models. Since this article focuses on data training for generative AI, it will not elaborate extensively on this classification.

3. Reconceptualizing fair use in the context of generative AI data training

China's Copyright Law draws on Articles 9 and 10 of the Berne Convention for the Protection of Literary and Artistic Works, formally introducing the "three-step test" to determine fair use. The "three-step test" establishes a framework for assessing fair use in copyright, which can only be recognized as compliant with fair use if the following three conditions are met: the use of a work should be in special circumstances, not conflict with the normal exploitation of the work, and not unreasonably prejudice the legitimate rights of the copyright holder. Applying the test to generative AI data training provides a theoretical basis for its legality.

3.1. Data training as a special circumstance

Due to the closed-list enumeration legislative model adopted in China's Copyright Law regarding fair use, the law explicitly lists twelve scenarios that can constitute fair use. However, in judicial practice, courts have deviated from these enumerated exceptions when confronting novel challenges arising from technological innovation and commercial development. In 2020, the Copyright Law underwent its third amendment, adding a catch-all clause (other circumstances as prescribed by laws or administrative regulations) in Article 24. This suggests that the "specific" and "particular" limitations under the "three-step test" should not be rigidly confined to the 12 statutory scenarios. Instead, even if the first step of the test is not satisfied, courts may still proceed to evaluate the second and third steps by synthesizing all relevant factors for a holistic determination [8].

As a key driving force in the digital economy era, AI data training provides a new model and space for human creativity, which is of great significance for promoting functional innovation and social development. It conforms to current technological trends and meets the requirements of predictability and purposefulness. Obviously, it passes the “first step” test. Therefore, it can be seen that the expansion interpretation of the catch-all clause of Article 24, Paragraph 13 can be made to

DOI: 10.54254/2753-7048/2025.BO23202

justify the fair use of the data training behavior of AI. The basis for this interpretation is that AI data training activities accord with the legislative purpose of Article 1 of the Copyright Law, which aims to promote the development and prosperity of socialist culture and science. However, in the long run, it is still necessary to determine the reasonable use of AI model training through special legislation and improve relevant laws to clarify the boundaries of this reasonable use [9].

3.2. Non-conflict with normal exploitation of works

According to the minutes of the meeting on amending the Berne Convention, "normal exploitation" is defined as "all uses capable of generating economic benefits or other potential pecuniary gains" [10]. From an economic perspective, "normal exploitation" can be conceptualized as the expected benefits derived from exercising copyright rights [11]. Thus, the crux of interpreting "normal exploitation" lies in whether the user's conduct competes with the copyright holder's market interests, thereby prejudicing the holder's conventional copyright-related gains. The impact of generative AI data training on the fair use of works can be explained from two aspects: technical operation mechanisms and the storage and presentation methods of works.

From the perspective of technical operation mechanisms, As previously discussed, AI training fundamentally constitutes machine learning----a process of generating new knowledge by analyzing existing data rather than reproducing works verbatim. At the input stage, copyrighted works are "fed" into the model, which uses deep learning to autonomously label data, identify latent correlations, and build algorithmic models. Upon receiving the new inputs, the model predicts outcomes based on these algorithms, resulting in the value-added knowledge creation rather than mere dissemination of the original works. This process does not compete with the original works or harm the rights holder's interests, thus preserving normal exploitation.

From the perspective of the storage and presentation of works, Generative AI training involves fragmenting and restructuring works into algorithmic parameters embedded within the model. It is worth noting that works are not stored intact or directly accessible in the model. This process is similar to human learning, where knowledge is applied "in blocks and fragments". From the perspective of copyright law, "reproduction" requires creating "one or more copies", yet AI training's transformative data processing does not meet this definition. Furthermore, since the work does not exist in its complete form in the model, the trained model lacks the conditions to become a market substitute for the work. Market substitutes typically have high similarity and substitutability with the original work in terms of content and function, while AI outputs are new creations distinct from training data in purpose and form. This ensures no interference with the original works' normal sales, distribution, or use.

3.3. No unreasonable prejudice to copyright holders' legitimate rights

The third step hinges on two key points: (1) it safeguards copyright holders' interests, not their rights; (2) "unreasonable" and "legitimate interests" imply that protected interests must be lawful, and harm to such interests must be unjustified. In essence, copyright holders bear a duty of tolerance for reasonable harm, while unlawful interests are excluded from protection. The impact of data training on this prong can be analyzed through governance costs and copyright holder interests.

From the perspective of governance costs, AI training relies on data with diverse and complex ownership structures. Requiring individual permissions for each work would render data collection operationally infeasible and incur prohibitive governance costs. Currently, no clear market mechanisms exist for large-scale work acquisition and transactions, leading to pricing difficulties and excessive transaction costs. Collective licensing schemes, while theoretically possible, would impose significant administrative burdens, with fees are likely to be absorbed by collective

DOI: 10.54254/2753-7048/2025.BO23202

management organizations, leaving copyright holders with disproportionately low returns [12]. Thus, considering both methodological appropriateness and teleological necessity, the use of works in generative AI training constitutes reasonable harm to copyright holders' interests, which means copyright holders have a certain duty of tolerance.

From the perspective of the impact on copyright holders, Generative AI training involves fragmenting works into statistical patterns stored as algorithmic parameters within the model. Outputs are neither direct reproductions nor derivative works of the training data, but rather belong to the technological market, distinct from the literary, artistic, and scientific domains [13]. This ensures no direct market competition with original works, preserving copyright holders' economic interests.

4. Dilemmas of the fair use system in the digital environment

While generative AI data training can theoretically satisfy the "three-step test" for fair use, its practical application reveals profound contradictions between legal institutions and technological advancements. The closed or semi-closed enumeration of fair use provisions in current law creates tensions with AI's open-ended data requirements, compounded by conflicts between rights-protection inclinations and the principle of technological neutrality, leading to ambiguous legal application.

4.1. The mismatch between existing fair use institutions and AI technology

China's Copyright Law Article 24 lists twelve scenarios of fair use, but these scenarios are difficult to match with the behavior of artificial intelligence data training. For instance, the first scenario, "using a work that has been published by others for personal study, research, or appreciation," is considered fair use, which applies only to individuals, excluding research teams, legal entities, or organizations. Artificial intelligence, as a systematic project, involves high research costs and complex tasks. Only relying on individual efforts is almost impossible to bear the related research expenses and complete the research independently [14]. Additionally, AI training often have commercial purposes, which does not match the non-commercial intent required under this provision. Another example is the sixth scenario, "translating, adapting, compiling, broadcasting, or making small-scale reproductions of published works for school teaching or scientific research purposes, provided that such works are not for publication or distribution," which is also considered fair use. Scientific research is limited to non-commercial institutions like public schools and research institutes. However, the subjects of AI data training not only include public research institutions but also involve a large number of private enterprises and commercial institutions, whose profit motives contradict the statute's non-commercial requirement. Furthermore, the "limited reproduction" requirement under this scenario differs from AI's need for massive datasets to improve models. Thus, it can be seen that the current regulations on the fair use system have not fully considered the situation of the rise of AI in the digital age. The existing system also cannot match the learning behavior of artificial intelligence in a perfect way.

4.2. Overemphasis on rights protection and its impact on fair use

Originating in late 18th-century France, "author-centricism" has a profound impact on the copyright law, driving continuous expansion of authors' rights in the digital age. At its core, this philosophy posits that works emanate from authors as extensions of their personality and spirit, entitling authors to comprehensive control over their creations [15]. Under this influence, the Berne Convention broadened the scope of reproduction rights to cover all known and unknown copying methods, making authorial control stricter. Taking digital libraries as an example, their digitalized

DOI: 10.54254/2753-7048/2025.BO23202

works may be regarded as an infringement of the right of reproduction, even though these actions are for the purpose of preserving and disseminating cultural resources and do not cause substantial harm to the interests of copyright holders. This tendency of overemphasizing the rights of the author has made the possibility of using works for reasonable purposes lower, and users face more legal risks when using works.

However, with the rise of artificial intelligence, "author-centricism" has gradually become weak. Creators are no longer exclusively human; AI-generated works are increasingly more as the main body of creation, challenging traditional authors' monopoly on creativity. As a key mechanism for ensuring "information access," fair use systems play an important role in the digital age. They provide legal ways for AI to learn from data and create, serving as crucial institutional support for the development of AI technology. But current copyright laws is still rooted in traditional models of creation, and provides excessive protection of the author's rights, failing to fully consider the needs of artificial intelligence's learning and leave enough space for the application of the reasonable use system in the field of artificial intelligence.

4.3. Disruption of the copyright interest balance

The core spirit of modern copyright legislation lies in coordinating the interests among copyright holders, work users, and the public to achieve a balance of interests. This spirit runs through the entire process of copyright law amendments and institutional design, serving as a critical guiding principle [16]. The emergence of artificial intelligence learning, however, has disrupted this balance.

The training of AI requires massive data, which often contains works protected by copyright law. During the process, developers are generally large internet enterprises that have overwhelming advantages in technology and financial resources, Thus weakening the control power of copyright holders over the works and threatening their expected interests [17]. Companies may use web crawlers and other data-mining tools to scrape copyrighted works without authorization, directly harming copyright holders and undermining the original balance. Moreover, if the current fair use system is directly applied to the data training, it is highly likely to result in an uneven distribution of benefits: Leveraging technological and resource advantages, AI generates more mature works after extensive training, obtaining greater benefits, while copyright holders cannot receive corresponding fair compensation. Broadly recognizing AI data training as fair use could suppress creative enthusiasm of authors, reducing work production and ultimately harming the public interest.

5. Reforms of the fair use system and mechanism innovations for the future

5.1. Shifting focus from rights protection to fair use

Although there is currently a fair use system in place, the entire copyright law system still demonstrates a tendency towards over - protecting the rights of copyright holders. The goal of regulating AI data training behavior is simply to protect the rights and interests of relevant parties. The key to rights protection lies in safeguarding the rights and interests enjoyed by all stakeholders in the training dataset, such as copyrights and personal information rights. It emphasizes the universality, indivisibility, and interdependence of rights. Essentially, its aim is to provide a baseline and uniform standard of protection for everyone, ensuring that the legitimate rights and interests of all parties are properly safeguarded [18]. However, with the advent of the digital age, the copyright system centered on "rights protection" has faced significant challenges. Excessive rights protection can make it difficult to obtain, use, and share AI training data information, leading to uneven distribution of training datasets and suppressing technological innovation and social benefits of generative AI.

DOI: 10.54254/2753-7048/2025.BO23202

The core of "fair use" lies in balancing interests and promoting the public interest. The fair use system, through restrictions on copyright, allows the public to legally use works under certain circumstances, avoiding the obstruction of knowledge dissemination and technological advancement due to over-protection of copyrights. Ultimately, it aims to promote the spread of knowledge, cultural prosperity, and technological innovation. Under the traditional framework of copyright law, works are regarded as extensions of the author's personality, with copyright being shaped into an exclusive absolute right. This "author-centric" assumption simplifies creative acts into isolated individual expressions while ignoring the essential characteristics of knowledge production-collectivity and accumulation. When AI technology achieves creative breakthroughs through massive data training, if the inevitable act of replicating work fragments during data training is uniformly regarded as infringement, it is tantamount to requiring all AI developers to negotiate licenses individually with each copyright holder. This is not only technically unfeasible but also stifles the possibility of technological innovation. Therefore, it is necessary to make "fair use" a goal of data training governance, reconstructing the existing regulatory framework to break the monopoly of "author-centricism".

5.2. Adding a dedicated AI fair use provision to balance innovation and copyright protection

While the "three-step test" can justify the fairness of generative AI data training, relying solely on teleological and expansive interpretations to fill legal gaps has significant limitations: the catch-all clause in Article 24 of China's Copyright Law remains overly vague, necessitating judges' discretionary interpretations of abstract concepts like "special circumstances" and "normal exploitation." This risks inconsistent judicial standards, particularly in generative AI, an area requiring high technological neutrality and massive datasets. Relying solely on interpretive approaches fails to provide enterprises with stable legal certainty and may lead to divergent rulings due to judges' differing understandings. Thus, theoretical justification is merely the first step; institutional implementation demands legislative clarity. To achieve this, a dedicated provision should be added (e.g. amending Article 24 to include "use of works for AI learning, training, and creation" as a fair use scenario) to demarcate the boundaries of AI data training and embed technological needs within the legal framework. Based on generative AI's operational mechanisms, work usage can be divided into input and output phases. Data training falls within the input phase, characterized by three key features under copyright law: (1) Massiveness: AI learning and training require vast quantities of copyrighted works as training data. (2) Fragmentation: AI processes works by decomposing and restructuring them into algorithmic parameters, rather than directly reproducing or copying them. (3) Educational Purpose: The use of works aims to enable internal learning and research within generative AI systems. As Mihály Ficsor observed, in principle, only public performances and public communications fall within copyright's scope [19]. Data training during the input phase does not involve public performances or communications, thus escaping copyright regulation and qualifying for fair use.

5.3. Breaking free from the 'closed' shackles: embracing openness and innovation

China's Copyright Law Article 24 adopts a "semi-closed" legislative model for fair use: the first twelve subparagraphs enumerate fixed scenarios, while the thirteenth----"other circumstances prescribed by laws or administrative regulations"----provides limited flexibility for emerging issues in judicial practice. However, this model has inherent limitations. The lack of uniform criteria for determining "other circumstances" undermines consistent application of fair use doctrines. Additionally, the "three-step test" suffers from ambiguous definitions of "special circumstances," unclear standards for "normal exploitation," and challenges in assessing unreasonable prejudice [20].

DOI: 10.54254/2753-7048/2025.BO23202

Thus, the new clause introduced in the 2020 amendment has failed to significantly facilitate fair use determinations, leaving China's fair use system trapped in a "closed" structure.

The key to overcoming this closure lies in shifting from "specific" to "general" standards, requiring both special-case provisions and general judgment criteria to establish a "specific-first, general-second" sequence. To generalize fair use, integrating the traditional "three-step test" with the four-factor test under U.S. Copyright Act Article 107 can create a clearer analytical framework. Specifically, the "step 2" (impact on normal exploitation) could incorporate the effect of use on the work's market value, while the "step 3" (unreasonable prejudice) could consider the purpose of use, the work's nature, the amount and quality of the portion used, as well as its relationship to the whole. The approach recognizes China's "trial-and-error" judicial practices, where courts fill legal gaps through flexible interpretation or legal creation in novel copyright disputes. Integrating these practices with the U.S. four-factor test provides a structured reference, enhancing the legitimacy of rulings. However, current applications of fair use doctrines in China lack rigor in selecting and reasoning about these factors, risking inconsistent outcomes. Therefore, retroactive recognition of "trial-and-error" practices must be conditional: courts must avoid imbalances in their experimental approaches to prevent judicial chaos.

6. Conclusion

In the era of digital civilization, generative AI data training challenges the traditional value system of copyright law. AI's utilization of works exhibits characteristics, functions, and effects fundamentally distinct from conventional copyright-recognized uses. Meanwhile, determining the copyright status of data training not only affects legal accuracy but also reshapes benefits distribution among stakeholders [21]. Through expansive interpretation, teleological analysis, and the "three-step test," generative AI's data training can be justified as fair use. However, AI's rapid evolution outpaces current legal frameworks, rendering existing fair use doctrines inadequate to support technological development. Mismatches between technology and institutions, combined with overemphasis on "rights protection," hinder digital progress. As German scholar Reinbothe aptly stated, "The essence of copyright law lies in balancing interests, not absolute protection." To address this, reforms and innovations in the fair use system are essential. By navigating the tension between "authorial personality rights" and "social communication rights," a new copyright law paradigm adapted to the digital age can emerge----one that serves as a dual engine for AI innovation and rights protection.

References

[1] Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC. (2019). Preamble, para. 8, and Article 2(2).
[2] Sobel, B. L. W. (2017). Artificial Intelligence's fair use crisis. Columbia Journal of Law & the Arts, 41(1), 45.
[3] Zhang, T. (2024). Legal risks of generative AI training datasets and inclusive prudential regulation. Journal of Comparative Law, (04), 86-103.
[4] Lin, X. (2021). Reshaping the fair use system in copyright law in the AI era. Chinese Journal of Law, 43(06), 170-185.
[5] Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. IBM Journal of Research and Development, 3(3), 210-211.
[6] Wei, Y. (n.d.). Copyright law response to generative artificial intelligence training data: Is it necessary to set fair use rules? Documentation, Information & Knowledge, 1-11.
[7] Wang, J. (2024). Research on copyright rules for artificial intelligence data training under the perspective of typology. Electronics Intellectual Property, (07), 20-30.
[8] Zhang, C. (2016). Misreading and Re-interpretation of "Three-Step Test" and "Fair-use Doctrine" in Chinese Copyright Law. Global Law Review, 38(05), 5-24.

DOI: 10.54254/2753-7048/2025.BO23202

[9] Xu, X. (2024). Fair Use of Copyright of Artificial Intelligence Model from Technology Neutrality Perspective. Law Review, 42(04), 86-99.
[10] World Intellectual Property Organization. (1971). Berne Convention for the Protection of Literary and Artistic Works. Revised on July 24, 1971.
[11] Xiong, Q. (2018). Judicial Standards for Copyright Fair Use: Resolving Ambiguities. Law Science, (01), 182-192.
[12] Samuelson, P. (2023). Fair Use Defenses in Disruptive Technology Cases. UCLA Law Review, 71, 1484. Retrieved March 15, 2024, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4631726
[13] Zhang, J., & Wang, S. (2024). Norms on Fair Use in Training Large Models. ECUPL Journal, 27(04), 20-33.
[14] Zhang, J. (2019). Fair Use of Artificial Intelligence: Dilemma and Solutions. Global Law Review, 41(03), 120-132.
[15] Liu, W. (2024). Philosophical Critique on Author-centrism of Authorship. Xiamen: Xiamen University Press.
[16] Feng, X. (2008). Research on the Theory of Balancing of Interests in Copyright Law. Journal of Hunan University (Social Sciences), 22(6), 113-120.
[17] Liu, Y., & Wei, Y. (2019). The Copyright infringement issue of Machine Learning and its Solution. ECUPL Journal, 22(02), 68-79.
[18] Zhang, T. (2024). A Risk-based Approach to AI Governance: Theory, Practice and Reflection. Journal of Huazhong University of Science and Technology (Social Science Edition), 38(02), 66-77.
[19] Fischer, M. (2009). The Law of Copyright and the Internet (Vol. 1). Encyclopedia of China Publishing House.
[20] Zhao, X. (2024). Fair Use of Generated Artificial Intelligence in Machine Learning. Jinan Journal (Philosophy & Social Sciences), 46(03), 79-95.
[21] Liu, X. (2024). "Non-Work Use" Nature of Generative Artificial Intelligence Data Training and Its Legitimization. Legal Forum, 39(03), 67-78.