

Article

Copyright Implications and Legal Responses to AI Training: A Chinese Perspective

Li You ¹ and Han Luo ^{2,*}

¹ Intellectual Property School, East China University of Political Science and Law, Shanghai 201620, China; 3249@ecupl.edu.cn

² School of Law, University of International Business and Economics, Beijing 100029, China

* Correspondence: 202200520111@uibe.edu.cn

Abstract

The emergence of generative AI presents complex challenges to existing copyright regimes, particularly concerning the large-scale use of copyrighted materials in model training. Legal disputes across jurisdictions highlight the urgent need for a balanced, principle-based framework that protects the rights of creators while fostering innovation. In China, a regulatory approach of “moderate leniency” has emerged—emphasizing control over downstream AI-generated content (AIGC) while adopting a more permissive stance toward upstream training. This model upholds the idea–expression dichotomy, rejecting theories such as “retained expression” or “retained style”, which improperly equate ideas with expressions. A critical legal distinction lies between real-time training, which is ephemeral and economically insignificant, and non-real-time training, which involves data retention and should be assessed under fair use test. A fair use exception specific to AI training is both timely and justified, provided it ensures equitable sharing of technological benefits and addresses AIGC’s potential substitutive impact on original works. Furthermore, technical processes like format conversion and machine translation do not infringe derivative rights, as they lack human creativity and expressive content. Even when training involves broader use, legitimacy may be established through the principle of technical necessity within the reproduction right framework.

Keywords: AI training; right of reproduction; derivative rights; copyright infringement; fair use



Academic Editor: Terry Carney

Received: 31 May 2025

Revised: 17 June 2025

Accepted: 18 June 2025

Published: 23 June 2025

Citation: You, Li, and Han Luo. 2025. Copyright Implications and Legal Responses to AI Training: A Chinese Perspective. *Laws* 14: 43. <https://doi.org/10.3390/laws14040043>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In recent years, generative artificial intelligence has witnessed explosive growth, with applications such as ChatGPT, Midjourney, Sora, Suno, and DeepSeek rapidly expanding in both reach and functionality. These technologies are reshaping the landscape of content creation, distribution, and consumption and have become key drivers of global technological innovation and industrial transformation. At the core of generative AI lies machine learning, where developers enhance the model’s capacity for learning and generalization through a process known as AI training. AI training, broadly construed, comprises two interrelated phases: data acquisition (including the collection, storage, and preprocessing of raw data) and data input (comprising pre-training, supervised fine-tuning, and reinforcement learning) (Sun and Dredze 2025). While the former constitutes a preparatory stage, the latter reflects the actual implementation of learning protocols and corresponds to what is typically referred to as AI training in the narrow sense. Unless otherwise indicated, this

article uses ‘AI training’ in its broader sense; however, in the analysis of specific practices, particularly in Section 3, the term is used in its narrow sense to refer exclusively to the model training process, excluding corpus construction or other preparatory activities.

The performance of generative AI systems depends heavily on the availability of large-scale, high-quality datasets. Unlike other types of AI, generative models must deeply analyze the language, stylistic features, compositional structures, and logical patterns embedded in intellectual creations across the literary, artistic, and scientific domains. Consequently, training datasets often contain copyrighted materials—such as books, artworks, music, and films—raising significant legal concerns regarding potential copyright infringement during the development process (Zurth 2021). These risks can be examined from two interrelated dimensions: the training output, which is a mathematical model generated through the learning of a dataset by machine learning algorithms, typically comprising a large number of parameters and a complex architecture, and the training process, which denotes the aggregate of activities performed by the developer or trainer in training a model, which involves the use of copyrighted materials as part of the training dataset. The former involves assessing whether the model has substantially retained the expressive elements of original copyrighted works; the latter centers on whether the acts involved in AI training amount to reproduction or the creation of derivative works under copyright law.

Globally, copyright litigation related to AI training has intensified across sectors such as publishing (Syndicat National de L’édition 2025), journalism (Canadian News Companies 2024), music (Sato 2024), movie (Shawn and Mat 2025), and visual arts (Vincent 2023). A pronounced conflict has emerged between copyright holders and generative AI developers (hereinafter referred to as “trainers”), with both sides disputing whether, and how, AI training activities should be regulated under existing copyright frameworks. In China, while there are no binding judicial decisions directly addressing copyright infringement claims involving AI training (Dong and Ren 2024), courts have begun exploring the copyrightability of AI-generated content (AIGC)¹ and the potential liability of generative AI service providers². On the regulatory front, the Interim Measures for the Management of Generative Artificial Intelligence Services (GAPM), the world’s first regulation specifically targeting generative AI, promulgated on 10 July 2023, and effective from 15 August 2023, explicitly enshrine the principle of “respect for intellectual property rights” in the context of AI training. These developments indicate that the interpretation and application of China’s current copyright regime will remain central to resolving future legal disputes in this evolving domain.

By contrast, in other jurisdictions, representative cases such as *Tremblay v. OpenAI*,³ *Silverman v. OpenAI*,⁴ and *Robert v. LAION*⁵ have substantively examined whether the AI training process falls within the scope of copyright protection. This interpretive, case-driven approach offers greater practical guidance for both theoretical development and judicial adjudication. In disputes concerning the ‘training output,’ the core issue is whether the process of training and AIGC constitutes unlawful copying of original works. In *Andersen v. Stability AI*, for instance, the plaintiff argued that the AI system compresses and stores the original images as ‘training images’ and that the output process constitutes a complex collage of preexisting works.⁶ This line of argument posits that the model essen-

¹ See Beijing Internet Court Civil Judgment (2023) Jing 0491 Minchu No. 11279; Changshu People’s Court of Jiangsu Province Civil Judgment (2024) Su 0581 Minchu No. 6697.

² See Guangzhou Internet Court Civil Judgment (2024) Yue 0192 Minchu No. 113.

³ See *Tremblay v. OpenAI, Inc.* 716 F.Supp.3d 772 (ND Cal 2024); *Tremblay v. OpenAI, Inc.* 742 F.Supp.3d 1054 (ND Cal 2024).

⁴ See *Silverman v. OpenAI, Inc.* F.Supp.3d (ND Cal 2023).

⁵ See *Robert Kneschke v LAION eV* (Hamburger Landgericht, Az. 310 O 227/23, 2024).

⁶ See *Sarah Andersen et al v Stability AI Ltd* 700 F Supp 3d 853 (ND Cal 2023).

tially functions as a ‘complex collage tool.’⁷ Infringement claims arising from the training process are typically more intricate and demand a careful distinction between two technical scenarios: ‘real-time training,’ in which copyrighted materials are temporarily loaded into memory for analysis, and ‘non-real-time training,’ in which data copies are fixed more permanently through mechanisms such as cloud storage or local hard drives. Furthermore, the training process often involves ancillary data-processing activities such as format conversion and language translation, adding further layers of legal complexity.

There is, as yet, no international consensus on how to address copyright infringement risks arising from both the training outputs and training processes of generative AI systems. Nevertheless, resolving the ongoing conflict between copyright holders and AI developers is crucial to unlocking the full technological and industrial potential of generative AI. Although the Copyright Law of the People’s Republic of China (the Copyright Law) does not explicitly address the applicability of copyright rights (Article 10 of the Copyright Law)⁸ or fair use categories (Article 24 of the Copyright Law)⁹ to AI training activities, developments in legislation and judicial practice reflect a clear policy orientation toward embracing technological advancement. In global academic and professional communities, three main theories have emerged as possible frameworks for resolving copyright risks associated with AI training: the non-expressive use theory, the fair use theory, and the statutory licensing theory.

The non-expressive use theory holds that the acts involved in AI training do not make use of the expressive elements of copyrighted works and therefore fall outside the scope of copyright regulation. The determinacy of the theoretical framework itself, especially its documented benefits for the development of small and medium-sized enterprises (SMEs) (Quang 2021), together with its cost-efficiency in practical implementation (X. Liu 2024), constitutes key evidence commonly relied upon by proponents to support their claims. The fair use theory asserts that the acts involved in AI training should or do constitute fair use and therefore do not amount to copyright infringement. Arguments in favor of including AI training within the scope of fair use often invoke doctrines such as the application of the transformative use standard (Sag 2019), the construction of a fair learning principle (Lemley and Casey 2020), adherence to the technology neutrality principle (X. Xu 2024), and a balancing of harms and benefits associated with use (Y. Liu 2024). These approaches are employed to justify both the legitimacy and necessity of recognizing AI training as falling within the ambit of fair use. The statutory license theory argues that the acts involved in AI training should be covered by a statutory license and therefore, after payment of compensation, do not constitute copyright infringement (Y. Cai 2024). It should be noted, however, that under current Chinese legislation, no statutory license exists that specifically applies to AI training, and the establishment of new regulations would be necessary to implement this approach. This is also true in most other countries. All three approaches rely on a precise legal delineation between the nature of the ‘training process’ and that of the ‘training output’.

The main goal of this article is to provide a solution to resolve the conflict between the development of artificial intelligence and copyright protection and to further enrich the body of doctrinal scholarship. While the copyright implications of AI training have been

⁷ See footnote 6 above.

⁸ The copyright rights most closely associated with AI training are the right of reproduction and derivative rights. Whether AI training activities constitute exercises of these rights remains legally unsettled and subject to interpretation. It is also noteworthy that Article 10, Item 17, of the Copyright Law contains a catch-all provision referred to as “other rights”, which may be invoked to extend protection to certain uses arising in the context of AI, depending on how such rights are interpreted and applied.

⁹ Article 24 of the Copyright Law enumerates twelve statutory exceptions constituting fair use and permits additional exceptions to be established by other laws or administrative regulations.

the subject of extensive scholarly discussion at the international level, the regulatory developments and practical approaches emerging from China have attracted comparatively little attention. Against this background, this article first adopts a China-focused perspective to examine the copyright risks of AI training, distilling recent developments and trends in domestic legal and policy practice. Second, by engaging with both domestic practice and comparative insights from other jurisdictions, this article analyzes the copyright implications arising from both the training process and training outputs, with particular focus on the legal characterization of acts involving the use of copyrighted works. Finally, this article aims to provide theoretical grounding and judicial guidance for a balanced resolution of this emerging legal challenge. It contends that whether AI training constitutes copyright infringement should turn on whether the relevant use of copyrighted content preserves the original expression protected under copyright law. This standard should apply uniformly to the acquisition, processing, and utilization of copyright-protected material throughout the AI training lifecycle.

2. Recent Developments in Chinese Practice

Since 2023, China has made significant strides in exploring the copyright issues related to AI training within its judicial and legislative practices. While there are no binding rulings or legal provisions that directly define the legal status of AI training, the ongoing developments suggest a clear trend: Chinese authorities have prioritized “post facto regulation” over “preemptive restriction”, signaling a proactive stance in fostering the growth of emerging technologies and industries.

2.1. Legislation: Focusing on Mitigating the Negative Externalities of AIGC

While relevant regulations have been introduced, China’s legislative control over AIGC is clearly more extensive than its approach to AI training, with the latter being treated as subordinate. Since 2022, China has enacted a series of prominent AI-related regulations, including Provisions on the Administration of Algorithm-generated Recommendations for Internet Information Services (ARR), Provisions on the Administration of Deep Synthesis of Internet-Based Information Services (DSR), GAPM, and the Measures for Identification of Artificial Intelligence-Generated Synthetic Contents (AIM). The common feature of these regulations is their focus on mitigating the negative externalities associated with AIGC while leaving the legal nature of AI training deliberately vague, either by not addressing it or by providing only limited discussion (Table 1).

The GAPM provide the most direct response to copyright risks associated with AI training and is closely linked to the subsequent Basic Requirements for the Safety of Generative Artificial Intelligence Services. Specifically, the former mandates that generative AI service providers must carry out pre-training, optimization, and other data processing activities in compliance with the law, using data and foundational models from legitimate sources. It also requires that these activities must not infringe upon the intellectual property rights of others.¹⁰ However, the GAPM does not address whether copyright holders have the right to control AI training, leaving this question unresolved and necessitating an examination of other legal sources.

The Basic Requirements for the Safety of Generative Artificial Intelligence Services, issued by the National Technical Committee 260 on Cybersecurity of the Standardization Administration of China, builds upon the GAPM. The general provisions of the Basic Security Requirements for Generative Artificial Intelligence Services state, “This document supports the GAPM and sets forth the basic security requirements that service providers

¹⁰ See Article 7 of the GAPM.

must follow. When service providers fulfill the filing requirements, they must conduct a security assessment in accordance with the provisions of Chapter 9 of this document and submit the assessment report”. Therefore, while the Basic Security Requirements for Generative Artificial Intelligence Services is not a typical administrative regulation or normative document in China, it still carries normative force.

Table 1. China’s latest AI-related regulations.

Name	Regulatory Target	Issued/Effective Date
ARR	Internet information services provided to users through algorithms such as generative synthesis, personalized recommendations, curated selections, search filtering, and scheduling decision-making technologies (Cyberspace Administration of China 2022a).	31 December 2021./1 March 2022.
DSR	Technologies that use generative synthesis algorithms—such as deep learning and virtual reality—to produce online content in the form of text, images, audio, video, and virtual environments (Cyberspace Administration of China 2022b).	25 November 2022./1 January 2023.
GAPM	Services that use generative artificial intelligence technologies to provide the public within the territory of the People’s Republic of China with content such as text, images, audio, and video (Cyberspace Administration of China 2023).	10 July 2023./15 August 2023.
AIM	Content is generated and synthesized through the use of artificial intelligence technologies (Cyberspace Administration of China 2025).	7 March 2025./1 September 2023.

2.2. *Judicial Practice: A Shift from Strict Liability to Moderate Leniency*

Although there are currently no binding precedents in Chinese judicial practice that directly address copyright infringement arising from AI training, two recent cases involving substantial similarity between AIGC and copyrighted works have begun to draw a connection between the actions of AI developers (or trainers) and potential copyright liability. While the courts differed in the degree of attribution, both decisions reflect an evolving judicial perspective that links training activities to infringement risks.

2.2.1. GIC Ultraman Case

On 8 February 2024, the Guangzhou Internet Court issued a decision in Shanghai Xinchuanghua Cultural Development Co., Ltd. v. Guangzhou Nianguang Co., Ltd., a case (GIC Ultraman case) that has been referred to by some media outlets as China’s first copyright infringement lawsuit involving an AI platform.¹¹ The plaintiff held the exclusive license to the copyright in the Ultraman series of artistic works within China and was entitled to enforce the rights independently. The defendant operated a website called “tab”, which provided generative AI services to the public via programmable APIs, accompanied by a paid subscription model. The plaintiff discovered that the AI system offered by the defendant could, upon user prompts, generate images that closely resembled those from the Ultraman series. Consequently, the plaintiff initiated legal proceedings, alleging that the defendant had infringed its rights of reproduction, adaptation, and communication to the public through information networks and sought to halt the infringement and demanded compensation for damages. The court framed the key issues of the case as whether the defendant’s actions infringed the plaintiff’s rights and, if so, how liability should be assigned.

The plaintiff argues that the defendant’s unauthorized generation of multiple Ultraman images infringes on the plaintiff’s right of reproduction over the Ultraman works

¹¹ See Guangzhou Internet Court Civil Judgment (2024) Yue 0192 Minchu No.113.

involved in the case. Some of the infringing images, such as those featuring Ultraman in an illustration style and the fusion of Ultraman with other characters like Sailor Moon and Doraemon, are substantially similar to the plaintiff's original works, thus infringing on the plaintiff's right of adaptation. Additionally, by generating and providing all the involved images to users, the defendant has violated the plaintiff's right of online communication of the Ultraman works.

The court ultimately concluded that the defendant was directly liable for infringing the plaintiff's rights of reproduction and adaptation. It also found the defendant at fault for failing to establish a complaint mechanism, provide risk warnings, or display prominent notices and therefore held the defendant liable for damages. The court ordered the defendant to implement an effective keyword filtering system to prevent the AI model from generating infringing content.

Regarding the right of reproduction, the court held that "the images involved in the case, generated by the tab website and provided by the plaintiff, partially or fully reproduce the original expression of the 'Ultraman' artistic character. Therefore, the defendant, without authorization, reproduced the Ultraman works, infringing on the plaintiff's right of reproduction over the Ultraman works involved". Concerning the right of adaptation, the court stated, "The generated images involved retain the original expression of the 'Tiga Ultraman Composite' work and, based on this original expression, develop new characteristics. The defendant's actions thus constitute an adaptation of the Ultraman works. Therefore, without permission, the defendant adapted the Ultraman works, infringing on the plaintiff's right of adaptation". Regarding the right of online communication, the court commented, "Whether the defendant infringed on the right of reproduction, adaptation, or the right of online communication is a matter of determining which specific copyright rights have been violated, and does not affect the establishment of infringement. In other words, it does not have a substantial impact on the interests of the rights holder or the public. Considering that this case involves a new situation of infringement arising in the context of generative AI, and given that the court has already upheld the plaintiff's claims of infringement on the right of reproduction and adaptation, the court will not re-evaluate the claim of infringement of the right of online communication as it is already covered under the claims for reproduction and adaptation". As for the plaintiff's request for the defendant to remove the Ultraman data from its training dataset, the court did not support this claim, as the defendant did not actually conduct model training.

The GIC Ultraman case sparked significant debate in both academic and industry circles upon its release, with some supporting the court's decision (Yao 2024) and others questioning its findings of fact and legal application (Juanmao 2024). The main issue arises from the court's reliance solely on the substantial similarity between the AIGC and the Ultraman series to conclude that the defendant had reproduced and adapted the plaintiff's copyrighted works without authorization (Luo and Yang 2025). However, the court did not clearly explain the relationship between the plaintiff's rights to reproduction and adaptation in the absence of direct involvement in the AI training process. In other words, it remains unclear whether the content generated by the model was an independent act of the defendant, the plaintiff, or the model itself or whether it was a collaborative result. The court's decision to find the defendant directly liable for infringement, rather than indirectly, reflects a tendency toward strict accountability. However, this perspective fails to explain why using third-party APIs to offer generative AI services constitutes acts of reproduction and adaptation. A more reasonable explanation is that the court did not adequately distinguish between scenarios in which the trainer directly provides services and those in which third parties indirectly provide services via APIs. Reproduction and adaptation, in this context, should be seen as the result of the actions taken during the development pro-

cess by the former. It should be noted that while this line of reasoning may explain the judiciary's perspective, it is not necessarily reasonable.

2.2.2. HIPC Ultraman Case

On 30 December 2024, the Hangzhou Intermediate People's Court issued its second-instance judgment in the case of Shanghai Character License Administrative Co., Ltd. v. Hangzhou Jellyfish Intelligent Technology Co., Ltd. (HIPC Ultraman case), affirming the first-instance decision.¹² This case marks China's second instance of AI platform-related copyright infringement litigation. In this case, the plaintiff, as the exclusive licensee of the Ultraman series of artistic works in China, had the right to enforce the copyright independently. The defendant operated a platform that utilized the open-source Stable Diffusion model, offering services for image generation, storage, distribution, and model customization under the framework of Stable Diffusion's text-to-image, image-to-image, and LoRA (Low-Rank Adaptation) technologies. The purpose of LoRA (Low-Rank Adaptation) technology is to allow users to upload a corpus of images and train a LoRA overlay model on top of a base model. This overlay model is used to supplement the content that the base model is unable to generate or to enhance specific image features, enabling fine-tuning and personalization of the generated images. This process helps tailor the images to better align with specific artistic styles and creative requirements. The plaintiff discovered that the defendant's platform not only allowed for the generation of images similar to those in the Ultraman series using a LoRA overlay model but also provided storage and distribution services for the generated content. As a result, the plaintiff initiated a lawsuit, claiming that the defendant had directly and indirectly infringed its rights to the communication of the Ultraman series works through information networks and engaged in unfair competition. The plaintiff sought to halt the infringement and demanded compensation for damages. The two courts summarized the key issues as "whether direct or indirect infringement occurred", "whether unfair competition was constituted", and "if infringement or unfair competition was found, how the defendant should be held liable". Ultimately, the courts concluded that the defendant should have been aware of the infringement but failed to take necessary action, thus constituting indirect infringement (specifically related to the right of communication via information networks). The courts ruled that the Anti-Unfair Competition Law would not be applied to further evaluate the case. The defendant was ordered to bear liability for damages, remove the infringing images, and eliminate any models capable of generating infringing content. It is noteworthy that when addressing the plaintiff's request to "delete all materials and data related to Ultraman", the first-instance court provided clarification on the legal boundaries of using copyrighted works in generative AI development. The court stated that 'in the absence of evidence showing that the generative AI was intended to use the original expression of a copyrighted work, or that it has affected the normal use of the work or unreasonably harmed the legitimate interests of the copyright holder, such actions may be considered fair use.'

Although the HIPC Ultraman case differs from the GIC Ultraman case in terms of the facts and the plaintiff's approach to enforcing their rights, it indirectly reflects a shift in Chinese judicial practice from strict liability to a more lenient stance concerning copyright risks in AI training. Specifically, the HIPC Ultraman case focuses more on the substitutive role of AIGC (artificial intelligence-generated content) with respect to copyrighted works—particularly its derivative effect in distribution—rather than on the legal nature of actions taken by parties during the development and application stages. As the plaintiff did not allege that the defendant had engaged in acts of reproduction or adaptation, the court did

¹² See Hangzhou Intermediate People's Court Civil Judgment (2024) Zhe 01 Minzhong No. 10332.

not address these aspects. However, the court in this case clarified the defendant's service provision based on an open-source model, thereby avoiding the confusion present in the GIC Ultraman case that mixed the scenarios of direct service provision by the trainer and indirect service provision via third-party interface calls. The court rejected the automatic assumption that AI service providers are directly infringing upon copyright. Based on this, some scholars have further argued the feasibility of applying the "safe harbor rule" to AI service provision scenarios and its positive implications for technological and industrial development (Xiong 2025). Additionally, the court differentiated the roles of the trainer, service provider, and user prior to the generation stage and introduced a framework for determining fair use based on the externalities of AIGC. This demonstrates that Chinese judicial practice is continuously evolving, with its reasoning becoming more refined, rational, and aligned closely with legislative logic.

2.2.3. BIC REDnote Case

Similar to other jurisdictions, China has not yet seen a definitive ruling directly addressing whether AI training constitutes copyright infringement. However, on 20 June 2024, the Beijing Internet Court began hearing a series of cases filed by four illustrators against REDnote (BIC REDnote case), revealing some trial details and providing a valuable window into the claims, demands, and evidence presented by both the plaintiffs and the defendant. The plaintiffs in these cases are illustrators who have long posted their artworks on the REDnote platform. They discovered that users had posted images on the platform that bore clear signs of imitation, with these users claiming that the images were generated using an AI drawing software provided by REDnote. This led to the dispute (Dong and Ren 2024). The plaintiffs assert that the defendant, without permission, used their artworks to train an AI model and applied it for commercial purposes, exceeding the boundaries of fair use and constituting copyright infringement. Specifically, the plaintiffs argue that the defendant's act of scraping their artworks to input them into the AI model infringes their right to reproduction; that the AI software's functionality, which combines their artworks with other images to create new works, infringes their right of adaptation; and that the defendant's actions also violate their right to control the use of their works as training material for AI. In defense, the defendant contends that there is no substantial similarity between the plaintiffs' artworks and the generated content, that the AI training constitutes fair use, and that they are not at fault. As of now, the case is still under trial.

In this case, the dispute between the plaintiff and the defendant centers around whether the "training result" and the "training process" lead to copyright infringement, an issue that is not explicitly addressed by China's current copyright legislation. On the one hand, the question arises as to whether the "training result" retains the copyrighted works and their adapted content as the original data, thereby infringing on the rights of reproduction and adaptation through the service-providing behavior. On the other hand, without the permission of the copyright holder, does the "training process" infringe on the rights of reproduction, adaptation, or other rights with a broad scope? Furthermore, can fair use provide an exemption for either of these two scenarios? The judiciary cannot refuse to render a judgment, so these issues remain to be clarified through judicial interpretation of China's current copyright law.

From the perspective of Chinese judicial practice, even if AI training does not fall within the statutory scope of the right of reproduction or the right of adaptation, the residual category of "other rights" under copyright law may still provide a legal basis for rights holders to pursue claims, provided certain conditions are satisfied. The Copyright Law establishes copyright powers in a relatively fragmented manner. In particular, the right of communication does not have a singular overarching provision like the "right

of making available to the public” but is instead divided into several specific rights, including the right of distribution, rental, exhibition, performance, screening, broadcasting, and online communication. With the in-depth development of technology and industry, new forms of work dissemination, such as online live streaming, have gradually become more widespread.

However, online live streaming, being a non-interactive transmission method via cable, does not fall under the scope of broadcasting rights (which cover wireless broadcasting, cable retransmission, and public performance broadcasts) as defined in the 2010 version of the Copyright Law, nor does it fall under the right of online communication (which covers interactive communication via wired or wireless means). The act of broadcasting works through online live streaming without authorization thus falls into a gray area between copyright infringement and non-infringement. Given the substantial impact such actions have on the rights holders, Chinese judicial authorities have interpreted the “other rights” provision to include online live streaming, addressing the copyright holders’ concerns.¹³ In the subsequent 2020 amendments, broadcasting rights were also expanded to cover this issue. It should be noted that the uncertainty surrounding the “other rights” provision and the risks associated with its broad application have sparked some criticism. The question of whether to retain this provision, and under what conditions it should be applied, remains a challenge for both legislators and the judiciary (Li 2022). Under the fair use doctrine, statutory exceptions such as “personal study or research” and “classroom teaching or scientific research” may also be analogically interpreted to exempt training activities (Wan and Li 2023). Whether such a doctrinal approach—relying on the existing copyright framework to regulate novel technological and industrial developments—represents a sustainable legal response remains open to scrutiny and evaluation.

3. “Training Outputs” and Copyright Infringement

As the tangible manifestation of AI model development, “training outputs” are the first point of contact for copyright holders and the broader public. Whether such outputs constitute copyright infringement depends on whether AIGC is similar to protected works at the level of ideas or at the level of expression—and, crucially, what causes that similarity. According to Article 9.2 of Agreement on Trade-Related Aspects of Intellectual Property Rights (the TRIPs Agreement), copyright protection extends only to expressions and not to ideas. Copyright law adheres to the idea–expression dichotomy, which protects only the specific form in which an idea is expressed, not the idea itself. For instance, if Person B rearticulates Person A’s academic viewpoint using different wording without attribution, this may constitute academic misconduct but not copyright infringement. Similarly, if an AI-generated painting resembles the style of a particular artist but differs significantly in its expression—such as in color composition, line work, and arrangement—then it would not constitute infringement. Only when the similarity extends beyond mere ideas and encompasses protected expressions would infringement potentially arise. Therefore, determining whether the AI model retains the expressive elements of the original training data is essential for assessing potential liability.

3.1. *Rejection of the Retained Expression Theory*

One line of reasoning contends that AI models—as the “output” of training—retain the expression of copyrighted works used as original training data. For instance, in *Anderesen v. Stability AI*, the plaintiff argued that the defendant’s “training conduct” resulted in the plaintiff’s works being stored in the AI model as “compressed copies”. Conse-

¹³ See Beijing Intellectual Property Court Civil Judgment (2017) Jing 73 Minzhong No.840.

quently, the AIGC produced—these “new” images—were claimed to be derivative works, and the AI-generated images were essentially complex “collages” of the plaintiff’s original artworks.¹⁴ Similarly, in the BIC REDnote case, the plaintiffs asserted that the AI services provided by the defendant could blend and mix their illustrations with other images to generate new ones, thereby constituting an infringement of the right of adaptation (Dong and Ren 2024). In essence, this argument also stems from the notion that the model retains the expressive elements of input data, thereby characterizing the service provider’s conduct as legally adaptive in nature.

To address the above concerns, it is necessary to clarify the technical principles underlying the practical process of AI training. At present, text-to-text (text generation) models and text-to-image (image generation) models are the mainstream forms of AI training models. The former aims to identify linguistic structures, syntax, and semantics in order to produce coherent text, while the latter is designed to recognize various objects, patterns, and features to generate images that align visually with user prompts. Despite their difference in output form, the two share the same underlying training logic: the model vectorizes input data and maps it into a frequency space, learning and adjusting its weights and biases—that is, its model parameters—to estimate the probability of the occurrence of relevant elements. Ultimately, what the model acquires is a statistical pattern of associations among elements rather than any wholesale or partial storage of the original training data for later retrieval. For example, when a user prompt includes the term “cloud”, the model can probabilistically predict related words such as “sky” or “raindrop”; similarly, when the term “secret of happiness” is input, the model may predict associated words like “health” or “wisdom” (Lohmann 2023).

The scope of copyright protection does not extend to associative relationships such as word frequency, syntactic patterns, or thematic markers, as doing so would substantially hinder the free dissemination of ideas. In fact, such associations are by no means novel in Chinese tradition. In the classical prosodic tradition, ancient texts exemplify the elegance of tonal and semantic symmetry. For instance, *Shenglü Qimeng* (《声律启蒙·上卷·一东》) contains the lines: ‘Morning to dusk, snow to frost in pair, River waves match lake’s mirrored glare. Pine winds echo bamboo’s tune, sunrise meets sunset.’ (朝对暮, 雪对霜, 江水对湖光, 松涛对竹韵, 旭日对夕阳). Likewise, *Liwen Duiyun* (《笠翁对韵·下卷·七阳》) presents: ‘Stillness to motion, chill to warmth, a lone boat to distant peaks, rosy clouds to misty haze, stars to the Milky Way.’ (静对动, 寒对温, 孤舟对远峰, 云霞对雾霭, 星斗对银河). In *Yinyun Jicheng* (《音韵集成·九青》), the verse reads: ‘A shepherd boy plays flute with mellow sound, field path and village swept by winds unbound. The boatman rows, the lake lies still and wide, where moonlight dances on the rippling tide.’ (牧童吹笛, 野陌穿村风袅袅。舟子摇橹, 平湖映月浪悠悠). These works demonstrate the artistry of phonetic and semantic correspondence. In lexicographical practice, the *Grand Chinese Dictionary* links over a hundred expressions to the character “water” (水), such as “water source” (水源), ‘truth emerges like water receding over rocks’ (水落石出), and “hydraulic engineering” (水利工程), revealing the richness and depth of Chinese morphological associations. Clearly, such linguistic or conceptual associations do not constitute copyright infringement of poems, essays, or books.

Accordingly, the model as a “training output” should not be classified as a copy or derivative work of the original data. As Judge Vince Chhabria of the U.S. District Court for the Northern District of California observed in *Kadrey v. Meta Platforms Inc.*:

“The plaintiffs allege that the ‘LLaMA language models are themselves infringing derivative works’ because the ‘models cannot function without the expressive information

¹⁴ See footnote 6 above.

extracted’ from the plaintiffs’ books. This is nonsensical. A derivative work is ‘a work based upon one or more preexisting works’ in any ‘form in which a work may be recast, transformed, or adapted.’ 17 U.S.C. § 101. There is no way to understand the LLaMA models themselves as a recasting or adaptation of any of the plaintiffs’ books”.¹⁵

In reality, what the model learns from the training dataset are high-frequency associations between texts and elements, rather than specific expressions (*Stability AI 2023*). For example, after training, the model can associate “gold” with “glitter”, but J.R.R. Tolkien’s famous phrase “All that is gold does not glitter” has little impact on the model’s weights compared to a less expressive sentence like “Gold glitters when exposed to light” (*U.S. Copyright Office 2023*). As the scale of the training dataset increases, the influence of any individual work diminishes. The model, in turn, learns generalized associative patterns that do not fall within the scope of copyright protection. This is conceptually aligned with the Merger Doctrine in copyright law, which withholds protection when an idea and its expression are indistinguishable due to the constraints of limited expressive forms.

Furthermore, post-training, models are unable to retain the works used as training data, nor do they replicate or store specific outputs (*Cooper et al. 2023*).¹⁶ AIGC is unlikely to closely resemble the original works, much like how a person might put a book down after reading it (*Lohmann 2023*). Experiments have shown that out of 175 million generated images, only 109 were near-duplicate training images. Accordingly, it can be established at least under current technological conditions that the model retains only a limited capacity for memorization (*Carlini et al. 2023*), and any resemblance in content is either coincidental or minimal. Technical measures, such as “deduplication”, are further reducing such instances. In the *Andersen v. Stability AI* case, the plaintiff failed to prove that the model’s output process replicated their works, and the court’s dismissal indirectly affirmed that models do not retain the expressions of the works used for training.¹⁷ In both the *GIC Ultraman* case and the *HIPC Ultraman* case, the courts determined that the liability to cease infringement could not, and should not, be achieved by deleting the original data or training materials from the model. Instead, the focus was placed on regulating the relationships at the output level, which similarly refutes the retained expression theory.¹⁸

3.2. Rejection of the Retained Style Theory

Another viewpoint asserts that AI models retains the style of the works, thereby constituting copyright infringement. For instance, in *Andersen v. Stability AI*, the plaintiff argued that their works possessed original stylistic elements, such as Sarah’s Scribbles, ‘combining classical realism, gothic and counterculture aesthetics’, and ‘a mixture of classical realism and impressionism’.¹⁹ However, style is considered an abstract idea, and copyright protection does not extend to ideas, a principle widely recognized in international legislation. Article 9(2) of the TRIPs Agreement explicitly states that Copyright protection shall extend to expressions and not to ideas, procedures, methods of operation or mathematical concepts as such. U.S. case law further reinforces this, recognizing that styles such as Impressionist or Abstract fall within the domain of ideas and thus are not copyrightable. As an example, a painter who develops a style or technique, such as the rendition of per-

¹⁵ See *Order Granting Motion to Dismiss, Kadrey v Meta Platforms, Inc.* Case No. 23-cv-03417-VC (TSH) (N.D. Cal. 2023).

¹⁶ A report from the International Conference on Machine Learning (ICML) points out that AI models, unlike search engines, cannot directly access their training data. Instead, they can only make predictions based on the information encoded in their model weights. The seminar lasts for two days, with the first day held as part of the International Conference on Machine Learning (ICML).

¹⁷ See footnote 6 above.

¹⁸ See Guangzhou Internet Court Civil Judgment (2024) Yue 0192 Minchu No.113; Hangzhou Intermediate People’s Court Civil Judgment (2024) Zhe 01 Minzhong No.10332.

¹⁹ See *Andersen v. Stability AI Ltd.* 744 F. Supp. 3d 956, 979 (N.D.Cal. 2024).

spective, impressionism, pointillism, fauve coloring, cubism, abstraction, psychedelic colors, minimalism, etc., cannot prevent others from adopting those ideas in their work.²⁰ From a legal perspective, extending copyright protection to style would undermine the fundamental principles of artistic creation. Although this approach may seem protective of the artist, it would ultimately restrict creative freedom.

Taking the Dutch painter Rembrandt as an example; he combined light-and-shadow techniques with profound depictions of human nature, pioneering new territories in Baroque painting. If style were protected, he would need to seek permission from earlier Italian masters like Caravaggio. Similarly, any artist who later adopted his light-and-shadow techniques would require authorization. While Rembrandt's influence spread across Europe and his techniques became the fashion of the 17th century, it is difficult to claim that he developed an entirely independent style. Nonetheless, this did not diminish his position in art history.

Consider the case of the Chinese calligrapher Wang Xizhi, renowned for his fluid running script and dynamic vitality, which has earned him a place in history. However, the initial strokes of his running script still bear traces of Han clerical script, and the essence of his work can be traced back to the aesthetics of the *Lanting Xu* (《兰亭序》). His bold brushwork may even have been influenced by Zhang Xu. These elements existed before Wang Xizhi, and some stelaes were still part of the tradition in his time. If style were restricted, would we then consider him to have plagiarized his predecessors? Similarly, if the Impressionist style of Monet or Leonardo da Vinci's anatomical realism were protected, subsequent artists would need permission to paint in these styles, effectively rendering artistic heritage a thing of the past. Countless such examples exist, as art creation generally begins with borrowing from the ideas of earlier thinkers. Every creation, to some extent, can be seen as originating from imitation, differing only in degree (Lu 2017).

Picasso holds copyright over his Cubist works, yet others are free to create within that style as long as they do not copy his specific expressions.²¹ Artistic progress depends on this reasonable inheritance. Protecting style alone would stifle innovation, which goes against the very essence of art.²²

Therefore, when discussing the potential copyright infringement of “training outputs”, it is important to recognize that algorithmic models learn statistical patterns of associations between elements within the works—the common templates present in the works—rather than the expressions protected by copyright. Aspects such as word frequency, sentence structure, themes, and other conceptual information fall outside the scope of copyright protection. Additionally, the fact that models are unlikely to retain specific expressions further supports this conclusion. This conclusion applies equally to the field of music.²³ Thus, “training outputs” retain only ideas, not expressions, and therefore do not constitute infringement. Furthermore, the fact that models do not preserve the expressions of the original works makes the discussion on whether the “training process” links the trainer to copyright infringement even more relevant and worthy of further examination.

4. “Training Process” and Copyright Infringement

When examining the issue of copyright infringement in the “training process”, the primary focus should be on whether the relevant actions meet the criteria for infringement of reproduction rights and derivative rights rather than immediately incorporating

²⁰ See *Jewelry 10, Inc v Elegance Trading Co.* No 88 Civ 1320 (PNL) (SDNY 1991).

²¹ See *Dave Grossman Designs, Inc. v Bortin* 347 F Supp 1150, 1156–57 (ND Ill 1972).

²² See *Nash v CBS, Inc.* 899 F 2d 1537, 1540 (7th Cir 1990).

²³ See *Arnstein v. Porter* 154 F.2d 464, 473 (2d Cir. 1946).

the discussion into the framework of fair use. This is because the Copyright Law frames discussions of rights limitations on the assumption that the right in question is already protected by copyright. In the absence of infringement (i.e., the presence of an actual right), limitations such as fair use or statutory licenses cannot be invoked.

4.1. “Training Process” and Infringement of Reproduction Rights

In analyzing reproduction rights, given the complexity of AI training, a blanket approach should not be applied. Instead, it is essential to consider the specific circumstances of the training process. The sources of data for trainers mainly include proprietary datasets, open-source datasets, publicly scraped or crawled datasets, and commercially purchased datasets (Tao 2024). Therefore, in industrial practice, data providers and trainers are not always the same entity. If the trainer does not establish a database and instead uses methods such as “cloud computing” and “federated learning” to feed training data into the model in real-time, these actions may be classified as “temporary reproductions” under copyright law (Zhu 2023). Due to their transient nature and lack of independent economic value, such actions are generally not subject to copyright regulation. However, if the trainer collects a significant number of works and stores them long-term on servers, this would certainly fall under reproduction rights, requiring further examination of whether such actions constitute “fair use”.

4.1.1. Real-Time Training: Justifying “Temporary Reproduction”

The first scenario in AI training involves real-time training. In this case, since the trainer does not directly acquire and permanently store works on their own servers, the temporary reproduction that occurs when the machine accesses the works in real-time raises concerns about potential copyright infringement. Some argue that “temporary reproduction” is still considered reproduction. In the context of AI training, when training materials are loaded into a computer’s memory (similar to how a work temporarily appears in memory when a webpage is accessed or files are viewed from a hard drive), this could potentially be seen as an infringement of reproduction rights (Lin 2021). Therefore, the key question becomes: does the reproduction involved in AI training qualify as “temporary reproduction”, and does it constitute an infringement of reproduction rights?

From a historical perspective, China has supported excluding “temporary reproduction” from the scope of reproduction rights in international negotiations. During the drafting of the WIPO Copyright Treaty (WCT), the inclusion of “temporary reproduction” within the scope of reproduction rights became a subject of debate. The treaty, aimed at regulating copyright protection in the digital environment, initially proposed in Article 7 of its “Basic Proposal” that “temporary reproduction” be included within the scope of reproduction rights. However, this was met with opposition from multiple country delegations (Mihály 2009). The Chinese delegation explicitly argued that reproduction rights should be limited to permanent reproduction only (World Intellectual Property Organization 1996). Mr. Shen Rengan, who was the Vice Administrator of the National Copyright Administration of China at the time and head of the Chinese delegation at the 1996 Geneva diplomatic conference organized by the WIPO for the formulation of the WIPO Copyright Treaty and the WIPO Performances and Phonograms Treaty, stated that the provision regulating temporary reproduction excessively expands the scope of protection for the right of reproduction, which he deemed unreasonable (Shen 1997; Shen and Zhong 2003). Due to these differing opinions, the diplomatic conference ultimately removed Article 7 from the proposal. WIPO Assistant Director General and representatives from multiple countries pointed out that the term “storage” could be interpreted by individual countries (Fraser

1997). This demonstrates China's stance favoring the exclusion of "temporary reproduction" and highlights the lack of international consensus on this issue.

Based on this, including "temporary reproduction" within the scope of reproduction rights is not an obligation that China must fulfill under international treaties. Article 31(2) of the *Vienna Convention on the Law of Treaties* stipulates that treaty interpretation can refer to any instrument which was made by one or more parties in connection with the conclusion of the treaty and accepted by the other parties as an instrument related to the treaty, but this declaration was not universally agreed upon. Also, the discussions at the conference confirmed that each country may independently define "storage". U.S. copyright scholar David Nimmer argues that "storage" refers to more stable, long-term actions that fall outside the scope of "temporary reproduction" and thus should not be included in the protection framework (Nimmer 1997). China's legislative practice also reflects this position. When drafting the Regulation on the Protection of the Right of Communication to the Public on Information Networks, there was a suggestion to regulate temporary reproduction. However, after careful consideration, this suggestion was not adopted (Ministry of Justice of the People's Republic of China 2006). The reasons were twofold: first, developing countries, including China, opposed the prohibition of "temporary reproduction" in the internet treaty (Ibid.); second, the Copyright Law does not authorize the regulation of "temporary reproduction", and subordinate legislation such as regulations shall not exceed the boundaries set by the law (J. Zhang 2006). This position aligns with China's stance during international negotiations.

From the perspective of purposive interpretation, the legitimacy of "temporary reproduction" at the doctrinal level can also be supported. The view that "temporary reproduction" should not be included within the scope of copyright can be factually substantiated from a technical standpoint. According to optical physics, when a digital device captures an image of a work, the entire technical process—from "lens imaging" to "memory caching" and finally to "screen display"—inherently results in "temporary reproduction". Even in the human observation phase, the biological temporary storage of the image on the retina constitutes a form of "temporary reproduction". While such physical processes may objectively enable permanent reproduction (e.g., through continuous observation leading to the reconstruction of memory or direct storage on a device), society generally accepts their legality.²⁴ The UK Supreme Court has clearly pointed out that denying the "temporary reproduction" exemption on the grounds that the user may not have actively terminated the device's operation represents an overextension and misinterpretation of the legislative intent.²⁵ In other words, the court distinguished between temporary and permanent reproduction, stressing that as long as the reproduction is part of a technical process and automatically concludes, any extension of the cache time by the user does not alter its temporary nature.

From a comparative law perspective, the legitimacy of the temporary reproduction exemption can also be supported. Legislation in countries such as Russia,²⁶ the Nether-

²⁴ Light is essentially an electromagnetic wave. When an image of a work is formed on the retina of the human eye, the rod and cone cells on the retina convert the light signals (electromagnetic waves) into electrical signals. These electrical signals are then transmitted to the occipital lobe, allowing the brain to process the image and colors. It can be said that when a person reads, the work is inevitably copied into the brain in an electronic form. This reasoning is undisputed from a scientific perspective. However, from the standpoint of general common sense, categorizing the temporary reproduction that occurs in this process as copyright infringement is undoubtedly absurd.

²⁵ See *Public Relations Consultants Association Ltd v Newspaper Licensing Agency Ltd and Others* [2013] UKSC 18, para 32.

²⁶ Article 1270, paragraph 2, of the гражданский кодекс (Civil Code of the Russian Federation, 2024 amendments) states: "As reproduction shall not be deemed a short term recording of a work which is of temporary or accidental nature and is an integral and significant part of a technological process solely intended for the

lands,²⁷ and Portugal²⁸ explicitly excludes “temporary reproduction” from the scope of control under reproduction rights. While Japan does not have a specific provision, both judicial precedents and scholarly consensus confirm the view that “temporary reproduction does not fall within the exclusive rights scope”.²⁹ Regulations in Taiwan of China and the Macao Special Administrative Region of China follows the same logic. Furthermore, European legal scholars, in their draft of the European Copyright Code, even extend the scope of “temporary reproduction” to all reproduction activities that lack independent economic value, thereby further solidifying the universality of this principle.³⁰

Chinese judicial practice strictly follows the statutory limitation that reproduction rights do not extend to “temporary reproduction”. For example, in the Yichawang case, the court determined that whether digital reproduction activities infringe upon reproduction rights must be assessed based on factors such as the duration of storage, technical necessity, and the independent economic value of the reproduced content (Ye and Sang 2017).³¹ The court in the Yichawang case explicitly denied the litigability of “temporary reproduction”, emphasizing that incorporating it into the scope of exclusive rights would contradict the legislative purpose of copyright law, which is to promote the flourishing of science and culture. Therefore, based on the threefold interpretation of statutory language, comparative legal consensus, and institutional values, “temporary reproduction” clearly falls outside the scope of regulation under reproduction rights under Chinese copyright regime.

This raises the question of whether real-time training constitutes a form of “temporary reproduction”. The substantive requirements for “temporary reproduction” include transient nature, technical necessity, and lack of independent economic value. In the context of real-time training, the act of loading works into memory satisfies the requirement of technical necessity, as it is an integral part of the technical process. The core debate, however, focuses on the proof of transient nature and the lack of independent economic value.

Firstly, from the perspective of the technical principles of real-time training, the physical properties of memory carriers inherently endow the process with the transient nature. The original training data, used for AI training, exist only temporarily in the computer’s memory and are automatically erased once the program terminates. They cannot be independently disseminated or reused outside the training system (Cooper et al. 2023). As such, even though the entire real-time training cycle may last for an extended period, the parsing of individual works in the neural network typically occurs in an instantaneous manner—similar to a round-the-world trip: while the overall journey may take a long time, the snapshot captured at each landmark during the trip is instant, triggered by the camera’s shutter.

Furthermore, as previously stated, AI training involves the disaggregation of the expressive content of training materials and the adjustment of weights to calculate probabilities. During AI training, only abstract features are extracted, rather than preserving the expression of the works themselves. To use the earlier analogy of photographing a

legal use of a work, or is the transfer of a work on an information telecommunication network between third parties by an information broker, provided that such record has no independent economic importance”.

²⁷ Auteurswet van 1912 (zoals gewijzigd tot 1 September 2017) Artikel 13a: Onder de verveelvoudiging van een werk van letterkunde, wetenschap of kunst wordt niet verstaan de tijdelijke reproductie die van voorbijgaande of incidentele aard is, en die een integraal en essentieel onderdeel vormt van een technisch procédé dat wordt toegepast met als enig doel (a) de doorgifte in een netwerk tussen derden door een tussenpersoon of (b) een rechtmatig gebruik van een werk mogelijk te maken, en die geen zelfstandige economische waarde bezit.

²⁸ See Código do Direito de Autor e dos Direitos Conexos (Code of Copyright and Related Rights, 2021amendments), Artigo 75.º, Âmbito 1.

²⁹ Article 21 The author shall have the exclusive right to reproduce his work. See *Japanese Copyright Act* (Act No. 48 of 1970, amended up to 19 July 2024).

³⁰ The right of reproduction is the right to reproduce the work in any manner or form, including temporary reproduction insofar as it has independent economic significance. Article 4.2 of European Copyright Code.

³¹ See Pudong New Area People’s Court of Shanghai Criminal Judgment (2015) Pudong Xing (Zhi) Chu No.12.

landmark, while the photo may capture the expression of another's architectural work, AI training does not preserve the expressive content of the work in the same way. The "automatic elimination" requirement emphasized by the European Court, although subject to theoretical debate, is inherently satisfied by deep learning models due to their technical nature.³² From an economic perspective, the fragmented storage of data and the algorithmic "black box" characteristics of the training process prevent the creation of independently tradable knowledge products,³³ thus lacking independent market value.

Given that real-time training constitutes a form of temporary reproduction under copyright doctrine, the non-expressive use theory has limited applicability in this context. Although both frameworks aim to exempt AI developers from liability, they diverge significantly in terms of their operative legal standards. The concept of temporary reproduction, while still recognizing the use as expressive in nature, adopts a broader and more technologically grounded perspective. It acknowledges the technical reality that real-time training frequently involves the ingestion of entire copyrighted works to optimize model performance. This approach is more doctrinally sound, as it avoids the impracticable task of isolating and proving which specific components of a protected work contribute meaningfully to the model's output—a burden that remains unfeasible given current technological constraints.

Moreover, the non-expressive use theory relies on an overly expansive standard, emphasizing technical or non-communicative uses (Quang 2021). However, this standard remains conceptually imprecise and analytically incomplete, as it cannot be meaningfully evaluated without reference to the downstream applications of AIGC. Even if the model itself does not retain elements of the original copyrighted work, it may nonetheless generate outputs that are identical or substantially similar to protected content, or that otherwise constitute expressive content. In such instances, relying on the non-expressive use theory to justify the legality of real-time training artificially severs the intrinsic connection between training and generation. This conceptual disjunction undermines the internal coherence of the theory and weakens its normative persuasiveness within the framework of copyright law (U.S. Copyright Office 2025).

4.1.2. Non-Real-Time Training: Exploring the "Fair Use" Approach

In addition to the "temporary reproduction" scenario, there exists a second situation where the trainer establishes a dedicated database on their server, using stored digital works and the associated training data formed through cleaning and annotation for training. This practice clearly differs from the aforementioned "temporary reproduction", as the storage activity meets the fixation requirement of reproduction under copyright law. As a result, this raises the issue of whether such behavior constitutes or should be regarded as fair use.

Article 24 of the Copyright Law does not explicitly address whether AI training can be subject to the fair use doctrine. However, even if one were to expand the interpretation or adopt a purposive approach to recognize AI training as fair use, there are still inherent limitations due to the "old wine in new bottles" problem, which prevents a comprehensive application of the doctrine. Even scholars who recognize the potential applicability of fair use to AI training have acknowledged the limitations of this approach and contend that legislative reform would offer a more coherent and sustainable solution (Wan 2021). The tradition of codified law exacerbates this issue. Nonetheless, despite the long revi-

³² See C-5/08 *Infopaq International A/S v Danske Dagblades Forening* [2009] ECR I-06569, paras 62–65.

³³ See Joined Cases C-403/08 and C-429/08 *Football Association Premier League Ltd and Others v QC Leisure and Others and Karen Murphy v Media Protection Services Ltd* [2011] ECR I-09083, paras. 174–177.

sion cycle of the Copyright Law,³⁴ the addition of the “catch-all provision” in the fair use system during the last round of amendments allows for new legal or administrative regulations to establish other fair use scenarios, thereby leaving room to address the legitimacy of AI training. According to explanations provided by Chinese legislators regarding the newly introduced open-ended fair use clause, its primary purpose is to accommodate the rapid development of the internet and to offer flexibility in addressing emerging scenarios. However, in order to prevent the abuse of judicial discretion, relevant authorities have also clarified that such a clause may only be applied in circumstances explicitly provided for by laws or administrative regulations and not be arbitrarily created by judicial bodies (Huang and Wang 2021). Therefore, although the HIPC Ultraman case explored the possibility of applying fair use to AI training, whether this reasoning can be effectively implemented in future cases remains uncertain in the absence of new legislative or regulatory provisions explicitly extending fair use to such contexts.

Should specific fair use provisions be established for AI training? To address this, it is necessary to deconstruct the issue through a three-step analysis. The most directly relevant question in the context of AI training pertains to the third step: “whether the legitimate interests of the copyright holder are unreasonably harmed”. This implies that even if harm is done to the legitimate interests of others, provided it falls within a reasonable scope, the fair use doctrine can still apply. Determining what is “reasonable” requires an examination from the perspective of legislative intent and value judgment. From one standpoint, the fair use doctrine has always been aligned with the utilitarian goals of copyright law. As such, there exists a natural balancing of interests between copyright limitations, exceptions, and the promotion of industrial development and societal progress. Fair use, with its inherent flexibility and openness, adapts to the evolving needs of society, fostering decisions that best serve public welfare.

On the other hand, from the perspective of value judgment, the public interest clearly outweighs the individual interests of copyright holders. While it is undeniable that AI represents a powerful competitor to humans, the widespread adoption and promotion of AI has, to some extent, displaced human authors in the market. However, the advancement of new technologies inherently leads to the obsolescence of older technologies, which aligns with the fundamental principles of social development and evolution. Technological advancement, by enhancing both the efficiency and quality of human creativity, has generated reciprocal benefits for authors and serves as a key rationale for reason why countries worldwide have implemented both explicit and implicit measures to “liberate” AI training. These policies and legislative trends reflect strategic national considerations—namely, the positive externalities of AI technology and industrial growth far surpass the individual interests of copyright holders. First, the technological benefits of AI are crucial to the international competitive landscape. Over-strengthening proprietary rights protection could result in the relocation of technological capital to jurisdictions with more lenient regulations, ultimately undermining the innovative potential of domestic enterprises (Sobel 2017). In contrast, establishing fair use rules would directly reduce the institutional costs associated with AI training. Second, the AI technological dividend is directly linked to the transformation of productivity. The significant role of AI in enhancing productivity is widely acknowledged, and this is built upon the advanced development of AI technologies. However, AI also presents a double-edged sword: its potential issues, such as hallucinations and bias, generate negative externalities that limit the full realization of its benefits. A comprehensive and thorough improvement in the authenticity, accuracy, objectivity, and diversity of training data is a common solution to achieving the former and

³⁴ The Copyright Law was promulgated and came into effect in 1990 and has since undergone three amendments in 2001, 2010, and 2020.

mitigating the latter. This, however, requires reasonable concessions in copyright protection, which serves as the institutional foundation for such progress (Ji et al. 2023). Thirdly, the extraction of AI technological dividends should aim at cost reduction and efficiency enhancement. The principle ‘entities should not be multiplied beyond necessity’ (*Numquam ponenda est pluralitas sine necessitate*) is equally applicable to the regulation of AI training by copyright law. If copyright regulation introduces unnecessary communication costs between trainers and copyright holders without meaningful progress, the rationale behind such measures becomes questionable. This is exemplified in the issue of statutory licensing. A substantial amount of evidence indicates that reaching a consensus between trainers and copyright holders is challenging (Statement on AI Training 2024). Even if new statutory licensing scenarios are introduced to address copyright holders’ concerns, key issues such as rate determination and mechanism construction will face considerable obstacles. The time costs involved will not align with the rapid pace of AI training, and the financial costs will further undermine the technological dividends of AI.

The adjustment of the fair use may serve as a viable solution to resolve the copyright legitimacy challenges posed by AI training. Examining the approaches of various countries’ copyright frameworks towards AI training, incorporating fair use represents a converging institutional innovation model, although the extent of its application differs across jurisdictions (Guan 2024). Given the recent legislative and judicial trends in China towards a more flexible approach, fair use is increasingly likely to become a mechanism for addressing the copyright risks associated with AI training. This could be achieved through the incorporation of targeted provisions into the revised Regulations for the Implementation of the Copyright Law, which are currently under amendment to reflect the 2020 Copyright Law revision and may potentially introduce a fair use exception applicable to AI training (L. Liu 2024). Parallel developments are expected under the forthcoming Artificial Intelligence Law, which has appeared in the State Council’s legislative agenda for both 2023 and 2024 (General Office of the State Council of the People’s Republic of China 2023, 2024).

To better engage copyright holders, it may be advisable to explicitly require that AI trainers share the technological benefits derived from AI development, such as by including fee reduction clauses. Furthermore, steps should be taken to avoid the generation of content that substitutes existing copyrighted material, which could be managed through mechanisms like “user instructions and output content controls” or the implementation of “input or output filters”. Finally, it is essential to retain the possibility for copyright holders to pursue accountability in cases where obligations are not fully fulfilled.

4.2. “Training Process” and Infringement of Derivative Rights

The issue of whether AI training infringes derivative rights is fundamentally linked to the question of reproduction rights infringement, as both rights serve analogous functions within the legal framework. Existing research often falls into the theoretical error of analyzing these two rights in isolation, which may lead to incorrect conclusions regarding AI training. One viewpoint posits that technical operations such as format conversion or language translation during AI training could infringe derivative rights (P. Zhang 2024). The flaw in this reasoning lies in a failure to recognize the inherent similarities between the regulatory functions of derivative rights and reproduction rights, both of which focus on the expression of the work.

In reality, derivative rights are only implicated when technical processes substantially alter the original work’s creative expression. When works are subjected to basic transformations, such as machine-readable encoding or language conversion, these processes remain fundamentally reproductions intended for the purpose of information processing and do not result in the creation of a new work possessing the requisite creative expression.

Thus, the current research paradigm fails to acknowledge the inherent connection between the rights' attributes, leading to an erroneous conclusion that purely technical modifications should be classified as derivative creations. A proper understanding of the scope of derivative rights requires distinguishing between "expression transformation" and "expression reconstruction". Only by making this distinction can we accurately assess the legal nature of data preprocessing activities in AI training and thereby establish a framework for infringement examination that is consistent with the technological characteristics of AI training.

The core premise for determining whether derivative rights are infringed is whether the act in question results in the creation of original expression. In the context of AI training, the various technical processes applied to digital works do not meet the originality standards set forth by copyright law, nor do they involve human creative intent (Wang 2023). As such, they do not constitute an infringement of derivative rights. The rationale for this lies in the fact that the digitization process (e.g., vector conversion of a painting) follows strict mathematical mapping rules, and the information conversion mechanism is governed by deterministic algorithms. Such mechanistic processing clearly cannot lead to the creation of original expression. Even when basic algorithms, such as high-frequency word selection, are employed, the restricted selection space makes it highly improbable that original expression will emerge. Moreover, even if the conversion process inadvertently results in a new form of expression, the process is entirely governed by pre-established algorithms, with no direct involvement of human creativity.³⁵ For instance, when a user employs translation software to mechanically convert a poem, the user merely executes operational commands without participating in the restructuring of expression. The resulting fixed translated text does not constitute derivative creation. Similarly, phenomena such as random events in nature (for example, a lightning strike altering the shape of a sculpture) further demonstrate that changes in expression driven by non-human forces cannot be attributed to human agency.

It is necessary to clarify that while technical processes do not constitute derivative works, they may still involve reproduction. Derivative rights govern the creation of new expression built upon the original, preserving its original creative elements while adding new expression. In contrast, reproduction rights regulate the mere reproduction of the original expression without introducing any new creativity. In the context of AI training, the technical processes involved in data preprocessing primarily function through "temporary reproduction" or "permanent reproduction" without independent economic value, serving the objective of information processing. As previously discussed, such reproduction, due to its technical necessity and value neutrality, can attain legitimacy under copyright law through a value-based assessment. Therefore, technical processes such as format conversion and language translation do not fall within the scope of derivative rights for secondary creation, nor do they involve substantial use prohibited by reproduction rights.

The value orientation of the copyright system, which seeks to incentivize creation, determines that derivative works generally exert a lesser negative impact compared to reproduction. A similar error can be found in U.S. judicial practice.³⁶ Some scholars have criticized this approach for its lack of coherence, arguing that it paradoxically results in non-original modifications escaping liability for copyright infringement, while original, creative modifications are penalized—an outcome that defies the principles of reason and fairness in copyright law (Cui 2014). This distinction provides an inherent basis for the legitimacy of derivative actions in AI training, beyond mere technical processes. In other

³⁵ See Article 3 of Regulation for the Implementation of the Copyright Law of the People's Republic of China (2013).

³⁶ See *Lee v A.R.T. Co* 125 F.3d 580 (7th Cir, 1997).

words, derivative actions and reproduction are intrinsically connected, and thus, the legality of derivative actions in AI training can be supported by the theoretical frameworks established for “temporary reproduction” and fair use.

On the one hand, due to the homogeneity of the underlying actions, even when AI training involves processing works that result in derivative works, such actions should at least be subjected to the same legal evaluation as reproduction actions. Since “temporary reproduction” and “permanent reproduction” without independent economic value have already attained legitimacy, derivative actions that preserve part of the original creative expression while adding new, independent creative expression should be exempt from infringement liability.

On the other hand, according to the legal principle of ‘weighing the heavier to clarify the lighter,’ derivative works, due to their creation of new originality, have a positive effect, and their negative impact is significantly lower than that of simple reproduction. Therefore, derivative actions should be afforded preferential treatment under the legal system. This stance is corroborated by China’s copyright law, particularly with respect to criminal and administrative responsibility. Article 217 of the Criminal Law of the People’s Republic of China regulates only reproduction and distribution, while Article 53 of the Copyright Law excludes most actions, including derivative works, from administrative liability. This reflects the legislator’s more lenient stance towards derivative actions.

5. Conclusions

The copyright risks associated with AI training are not only a concern for China but also a global challenge, which is confronted with the dual pressures of technological innovation and regulatory reasoning. Jurisdictions are actively formulating responses to AI-related copyright challenges in light of their distinct economic, political, cultural, and technological conditions. For instance, the United States has adopted a posture of cautious intervention, asserting that the doctrine of fair use is capable of resolving the majority of legal issues arising from AI training, while residual concerns may be addressed through self-correcting market mechanisms ([U.S. Copyright Office 2025](#)).

Similarly, the European Union, through the Copyright in the Single Market Directive, has established a framework of exceptions for text and data mining (TDM), accompanied by an opt-out mechanism. While these provisions offer a potential pathway for reconciling the interests of AI development and copyright protection, their practical application remains subject to considerable uncertainty, particularly in commercial contexts involving AI training ([European Union Intellectual Property Office \(EUIPO\) \(2025\)](#)). The Japanese Copyright Act has likewise introduced a provision permitting non-enjoyment uses, which has substantially lowered the legal barriers to AI training ([Agency for Cultural Affairs 2018](#)). Nonetheless, this legislative development has been met with sustained opposition from copyright holders ([Deck 2023](#)). The ongoing divergence in regulatory approaches highlights the absence of a broadly accepted consensus on how to reconcile AI training with copyright protection.

In China, the latest AI-related regulations, including ARR, DSR, GAPM, and AIM, establish a governance framework focused on preemptively curbing and subsequently mitigating the negative externalities of AIGC. The ambiguity in the legal treatment of AI training reflects a policy that encourages technological and industrial development, emphasizing inclusive growth. This approach has also been progressively refined and consolidated through relevant judicial practice. These approaches provide a pragmatically oriented framework for addressing the copyright implications of AI training.

This article argues that the determination of copyright infringement should focus on whether the use of a copyrighted work retains its original expression, with AI training’s

acquisition, processing, and utilization of copyrighted works being subject to this criterion. At the “training output” level, based on the technical principles of AI training, the AI model itself does not serve as a repository for copyrighted works. If the AIGC does not retain the original expression of the copyrighted work, the trainer, being involved only in non-protectable elements such as ideas, styles, or facts, does not constitute copyright infringement. Consequently, both the retained expression theory and retained style theory are untenable.

At the “training process” level, it is critical to distinguish technical scenarios involving the digital storage of copyrighted works. The transient nature and lack of independent economic value inherent in “real-time training” naturally categorize it within the scope of “temporary reproduction”, whereas “non-real-time training”, which involves the creation of a database and the formation of stable copies, requires a fair use review. The homogeneity of derivative rights and reproduction rights implies that algorithm-driven technical processes such as work format conversion and language translation, which lack human creative intent and the generation of original expression, do not breach the boundaries of derivative rights. Even outside technical processing, any derivative actions undertaken by the trainer on copyrighted works can attain legitimacy under the framework of reproduction rights, provided they meet the principle of technical necessity.

Author Contributions: Conceptualization, L.Y. and H.L.; Writing—original draft, L.Y.; Writing—review & editing, L.Y. and H.L.; Visualization, L.Y. and H.L.; Supervision, L.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Agency for Cultural Affairs. 2018. Overview of the Amendment to the Copyright Act. Available online: https://www.bunka.go.jp/seisaku/chosakuken/hokaisei/h30_hokaisei/pdf/r1406693_02.pdf (accessed on 15 June 2025).
- Cai, Yuanzhen. 2024. Copyright Statutory Licensing of Machine Learning: Foundations and Regulations. *Intellectual Property* 38: 77.
- Canadian News Companies. 2024. Challenge OpenAI over Alleged Copyright Breaches. CNBC, November 29. Available online: <https://www.cnn.com/2024/11/29/major-canadian-news-media-companies-launch-legal-action-against-openai.html> (accessed on 1 January 2025).
- Carlini, Nicolas, Jamie Hayes, Milad Nasr, Matthew Jagielski, Vikash Sehwal, Florian Tramèr, Borja Balle, Daphne Ippolito, and Eric Wallace. 2023. Extracting Training Data from Diffusion Models. pp. 5–6. Available online: <https://arxiv.org/pdf/2301.13188.pdf> (accessed on 5 May 2025).
- Cooper, A. Feder, Katherine Lee, James Grimmelmänn, Daphne Ippolito, Christopher Callison-Burch, Christopher A. Choquette-Choo, Niloofar Miresghallah, Miles Brundage, David Mimno, Madiha Zahrah Choksi, and et al. 2023. Report of the 1st Workshop on Generative AI and Law. p. 27. Available online: <https://arxiv.org/abs/2311.06477> (accessed on 5 May 2025).
- Cui, Guobin. 2014. *Copyright Law: Cases and Materials*. Beijing: Peking University Press, p. 419.
- Cyberspace Administration of China. 2022a. ARR, Q&A. January 4. Available online: https://www.cac.gov.cn/2022-01/04/c_1642894606594726.htm (accessed on 23 March 2025).
- Cyberspace Administration of China. 2022b. DSR, Q&A. December 12. Available online: https://www.gov.cn/zhengce/2022-12/12/content_5731430.htm (accessed on 23 March 2025).
- Cyberspace Administration of China. 2023. Official Q&A on the GAPM. July 15. Available online: https://www.gov.cn/zhengce/2023-07/content_6892001.htm (accessed on 23 March 2025).
- Cyberspace Administration of China. 2025. Q&A on the AIM. March 14. Available online: https://www.cac.gov.cn/2025-03/14/c_1743654685896173.htm (accessed on 23 March 2025).

- Deck, Andrew. 2023. This Japanese Manga Artist-Turned-Politician Is Taking on AI Art. *Rest of World*, February 16. Available online: <https://restofworld.org/2023/generative-ai-japanese-politicians-manga/> (accessed on 16 June 2025).
- Dong, Wenjia, and Huiying Ren. 2024. Beijing Internet Court Trials First Case of Copyright Infringement Involving AI Painting Model Training. Beijing Internet Court WeChat Official Account. June 20. Available online: <https://mp.weixin.qq.com/s/cyskAz1cASBaNIYQpGpGsA> (accessed on 1 January 2025).
- European Union Intellectual Property Office (EUIPO). 2025. Development of Generative Artificial Intelligence from a Copyright Perspective. May 12. Available online: https://euipo.europa.eu/tunnel-web/secure/webdav/guest/document_library/observatory/documents/reports/2025_GenAI_from_copyright_perspective/2025_GenAI_from_copyright_perspective_FullR_en.pdf (accessed on 15 June 2025).
- Fraser, Stephen. 1997. The Copyright Battle: Emerging International Rules and Roadblocks on the Global Information Infrastructure. *The John Marshall Journal of Computer & Information Law* 15: 777.
- General Office of the State Council of the People's Republic of China. 2023. Notice on the Issuance of the State Council's 2023 Legislative Work Plan. June 6. Available online: https://www.gov.cn/gongbao/2023/issue_10526/202306/content_6887136.html (accessed on 4 April 2025).
- General Office of the State Council of the People's Republic of China. 2024. Notice on the Issuance of the State Council's 2024 Legislative Work Plan. May 9. Available online: https://www.gov.cn/zhengce/content/202405/content_6950093.htm (accessed on 4 April 2025).
- Guan, Chunyan. 2024. Exploring the Fair Use of Copyright for Generative Artificial Intelligence Training: International Trends, Local Development, and the Construction of Rules. *Publishing Research* 40: 91.
- Huang, Wei, and Leiming Wang, eds. 2021. *Introduction and Interpretation of the Copyright Law of the People's Republic of China*. Beijing: China Democratic and Legal Publishing House, pp. 154–55.
- Ji, Jessica, Josh Goldstein, and Andrew Lohn. 2023. Controlling Large Language Models: A Primer. Center for Security and Emerging Technology. December. Available online: <https://cset.georgetown.edu/wp-content/uploads/CSET-Controlling-Large-Language-Model-Outputs-A-Primer.pdf> (accessed on 5 May 2025).
- Juanmao. 2024. First Global AIGC Platform Infringement Case Decided: “Ultraman” Defeats AI. March 4. Available online: <https://news.qq.com/rain/a/20240304A03D2Y00> (accessed on 5 April 2025).
- Lemley, Mark A., and Bryan Casey. 2020. Fair Learning. *Texas Law Review* 99: 743–79. [CrossRef]
- Li, Chen. 2022. On the Interpretation of Other Rights which Shall be Enjoyed by the Copyright Owners. *Intellectual Property* 36: 21.
- Lin, Xiuqin. 2021. Reshaping the Fair Use System in Copyright Law in the AI Era. *Chinese Journal of Law* 43: 176.
- Liu, Liang. 2024. In the First Nine Months of the 7th China International Import Expo, over 30,000 Trademark Infringement and Fake Patent Cases Were Handled in China. *China News Network*, November 6. Available online: <https://www.chinanews.com.cn/cj/2024/11-06/10314706.shtml> (accessed on 4 April 2025).
- Liu, Xiaochun. 2024. Non-Work Use Nature of Generative Artificial Intelligence Data Training and its Legitimization. *Legal Forum* 39: 67.
- Liu, Yu. 2024. An Economic Analysis of Machine Data Utilization Constituting Fair Use under Copyright Law. *Intellectual Property* 38: 107.
- Lohmann, Fred. 2023. Re: Notice of Inquiry and Request for Comment [Docket No. 2023-06]. October 30, pp. 5–6. Available online: https://downloads.regulations.gov/COLC-2023-0006-8906/attachment_1.pdf (accessed on 5 May 2025).
- Lu, Haijun. 2017. On the Nature of Idea/Expression Dichotomy. *Intellectual Property* 31: 25.
- Luo, Han, and Ruiqi Yang. 2025. Reflection and Exploration on Identification of Copyright Infringement in Artificial Intelligence Generated Content—A Case Study of the Ultraman. *Digital Publishing Research* 4: 69.
- Mihály, Ficsor. 2009. *The Law of Copyright and the Internet*. Translated by Guo Shoukang, Wan Yong, and Xiang Jing. Beijing: China Encyclopedia Publishing House, vol. 1, pp. 178–98.
- Ministry of Justice of the People's Republic of China. 2006. The Head of Legislative Affairs Office of the State Council of the People's Republic of China Responds to the Question from China Government Legal Information Network's Reporter Regarding the 'Regulation on the Protection of the Right of Communication to the Public on Information Networks'. December 4. Available online: https://www.moj.gov.cn/pub/sfbgw/zcjd/200612/t20061204_389820.html (accessed on 19 December 2024).
- Nimmer, David. 1997. A Tale of Two Treaties-Dateline: Geneva—December 1996. *Columbia-VLA Journal of Law & the Arts* 22: 15–16.
- Quang, Jenny. 2021. Does Training AI Violate Copyright Law? *Berkeley Technology Law Journal* 36: 1420–29.
- Sag, Matthew. 2019. The New Legal Landscape for Text Mining and Machine Learning. *Journal of the Copyright Society of the USA* 66: 291–319. [CrossRef]
- Sato, Mia. 2024. Major Record Labels Sue AI Company behind 'BBL Drizzy'. *The Verge*, June 24. Available online: <https://www.theverge.com/2024/6/24/24184710/riaa-ai-lawsuit-suno-udio-copyright-umg-sony-warner> (accessed on 1 January 2025).

- Shawn, Chen, and O' Brien Mat. 2025. Disney and Universal Sue AI firm Midjourney for Copyright Infringement. *AP News*, June 12. Available online: <https://apnews.com/article/disney-universal-midjourney-copyright-lawsuit-722b1b892192e7e1628f7ae5da8cc427> (accessed on 15 June 2025).
- Shen, Rengan. 1997. WIPO Introduced Two New Treaties. In *Intellectual Property Research*. Edited by Zheng Chengsi. Beijing: China Fangzheng Press, vol. 3, p. 9.
- Shen, Rengan, and Yingke Zhong. 2003. *Introduction to Copyright Law*, rev. ed. Beijing: The Commercial Press, pp. 244–46.
- Sobel, Benjamin L. W. 2017. Artificial Intelligence's Fair Use Crisis. *Columbia Journal of Law & the Arts* 41: 81.
- Stability AI. 2023. Response to United States Copyright Office Inquiry into Artificial Intelligence and Copyright. October, p. 13. Available online: https://www.law.berkeley.edu/wp-content/uploads/2023/11/Stability-AI-COLC-2023-0006-8664_attachment_1.pdf (accessed on 5 May 2025).
- Statement on AI Training. 2024. Available online: <https://www.aitrainingstatement.org/> (accessed on 4 April 2025).
- Sun, Kaiser, and Mark Dredze. 2025. Amuro and Char: Analyzing the Relationship Between Pre-Training and Fine-Tuning of Large Language Models. March 18. Available online: <https://arxiv.org/pdf/2408.06663> (accessed on 15 June 2025).
- Syndicat National de L'édition. 2025. Authors and Publishers Unite in Lawsuit Against Meta to Protect Copyright from Infringement by Generative AI Developers. March 18. Available online: <https://www.sne.fr/press-release-authors-and-publishers-unite-in-lawsuit-against-meta-to-protect-copyright-from-infringement-by-generative-ai-developers/> (accessed on 23 March 2025).
- Tao, Qian. 2024. Copyright Problems regarding Training of Foundation Models: Clarification of the Theory and Application of Rules. *Tribune of Political Science and Law* 42: 152.
- U.S. Copyright Office. 2023. Notice of Inquiry on Artificial Intelligence & Copyright (Dkt. 2023–2026) Reply Comments of Meta Platforms. Inc. December 6, p. 13. Available online: https://downloads.regulations.gov/COLC-2023-0006-10332/attachment_1.pdf (accessed on 5 May 2025).
- U.S. Copyright Office. 2025. Copyright and Artificial Intelligence, Part 3: Generative AI Training (Pre-Publication Version). May, pp. 47–48. Available online: <https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-3-Generative-AI-Training-Report-Pre-Publication-Version.pdf> (accessed on 15 June 2025).
- Vincent, James. 2023. Getty Images Sues AI Art Generator Stable Diffusion in the US for Copyright Infringement. *The Verge*, February 7. Available online: <https://www.theverge.com/2023/2/6/23587393/ai-art-copyright-lawsuit-getty-images-stable-diffusion> (accessed on 1 January 2025).
- Wan, Yong. 2021. The Dilemma and Solution of the Fair Use System under Copyright Law in the Age of Artificial Intelligence. *Social Sciences Journal* 43: 93.
- Wan, Yong, and Yalan Li. 2023. Research on the Interpretation of Fair Use Clause in Response to the Development of Artificial Intelligence Industry. *Digital Law* 1: 83.
- Wang, Qian. 2023. The Qualitative Analysis of Content Generated by Artificial Intelligence in Copyright Law. *Tribune of Political Science and Law* 41: 24.
- World Intellectual Property Organization. 1996. Amendments to Partly Consolidated Text of Draft Treaty No. 1. CRNR/DC/64. *Delegation of the Peoples Republic of China*, December 13.
- Xiong, Qi. 2025. Copyright Infringement Liability of Generative Artificial Intelligence Platforms. *Global Law Review* 47: 23.
- Xu, Xiaoben. 2024. Fair Use of Copyright of Artificial Intelligence Model from Technology Neutrality Perspective. *Law Review* 42: 86.
- Yao, Zhiwei. 2024. Determination and Prevention of Copyright Infringement of AI Generated Works: Focusing on the Worlds First Generative AI Service Infringement Judgment. *Local Legislation Journal* 9: 1.
- Ye, Jufen, and Qingyuan Sang. 2017. Storing Transcoded Novels on Webpages Constitutes Infringement. *Peoples Court Daily (Beijing)*.
- Zhang, Jianhua, ed. 2006. *Interpretation of Regulation on the Protection of the Right of Communication to the Public on Information Networks*. Beijing: China Legal Publishing House, p. 5.
- Zhang, Ping. 2024. The Obstacles and Solutions of Copyright System in Artificial Intelligence Content Generation Mechanism. *Science of Law* 42: 27.
- Zhu, Kaixin. 2023. Understanding the Core Copyright Issues in Training Large AI Models. *Tencent Research Institute*, October 19. Available online: <https://mp.weixin.qq.com/s/ab1p8v2QopyCmHXzdkcQmg> (accessed on 6 April 2025).
- Zurth, Patrick. 2021. Artificial Creativity? A Case Against Copyright Protection for AI-Generated Works. *UCLA Journal of Law and Technology* 25: 18.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.