Proceedings of ICILLP 2025 Symposium: Property Law and Blockchain Applications in International Law and Legal Policy
DOI: 10.54254/2753-7048/2025.LD28196

A Study on the Determination and Allocation of Liability for Copyright Infringement by Generative Artificial Intelligence

Yueyi Ma

Dalian University of Technology, Dalian, China
myy19043345282@qq.com

Abstract. With the rapid development of generative artificial intelligence (generative AI) technologies, while these technologies have greatly facilitated creative production, they have also spawned numerous challenges for determining and allocating liability for copyright infringement. This article first analyzes the technical principles and characteristics of generative AI, identifies data as the critical factor in infringement determinations, and clarifies core elements such as the standards for attributing copyright infringement liability, while highlighting difficulties—among them the complex and multifaceted nature of potential liable parties in AI-related infringements. Thereafter, from both legislative and judicial-practice perspectives, the paper examines current approaches: legislatively, China relies on the Copyright Law and the Civil Code as its basic framework; in judicial practice, it analyzes cases such as the Hangzhou "Ultraman" case and the Guangzhou AI text-to-image case and, through comparative analysis with foreign jurisdictions, addresses the above problems. The article advances multi-dimensional recommendations: at the legislative level, clarify ownership and rights in AI-generated content; at the technical and industry level, promote the adoption of "data whitelists" and provenance/traceability technologies; and at the international level, strengthen cooperation to align with international norms and to resolve cross-border disputes — with the ultimate goal of achieving a dynamic balance between legal regulation and technological development.

Keywords: generative artificial intelligence, copyright infringement, liability determination, liability allocation

1. Introduction

1.1. Research background of copyright infringement by generative artificial intelligence

The rise of generative artificial intelligence is undoubtedly a major trend of our era: it offers convenience to people across industries and demonstrates powerful content-creation capabilities. However, the works produced by generative AI face many challenges with respect to the determination and protection of copyright.

Under current international legal frameworks, existing copyright laws are primarily designed with human authors as the rights-holders, and clear legal rules have yet to be formed concerning questions such as ownership of works created by generative AI and the boundaries of related rights

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-7048/2025.LD28196

[1]. Moreover, significant controversies remain in the concrete application of the law and the drafting of rules to address these issues.

Recently, the Thomson Reuters case has attracted wide attention. That case constitutes, to date, the first substantive U.S. judgment addressing whether the training of large AI models constitutes copyright infringement. Coming before further amendments to U.S. fair-use legislation and before the U.S. Copyright Office issued guidance on fair use in the AI context, the decision again underscored how U.S. courts may affect the development of the American AI industry. Domestically, with the widespread application of generative AI, related copyright disputes have gradually increased. In November 2023, the Beijing Internet Court's decision in an AI text-to-image case was regarded as China's first landmark AI copyright case. Therefore, an in-depth study of how to determine and allocate liability for copyright infringement by generative AI is of significant practical importance for regulating the healthy development of the AI industry and for protecting the legitimate rights and interests of creators. This paper aims to analyze the attribution of copyright infringement liability for AI-generated works by means of case investigations and case-by-case analysis.

Against the backdrop of rapidly advancing digital technologies, generative AI is transforming creative models, and the problems of determining and allocating copyright liability urgently need resolution. Research on this topic carries important practical significance. Traditional theories contain many gaps when confronted with AI-generated content; in-depth study can help fill these theoretical voids in copyright liability for new technologies and provide a more forward-looking theoretical foundation for subsequent research, enabling the legal framework to better adapt to the new challenges posed by technological progress. Because clear rules for determination and allocation are lacking, judges face considerable difficulties when handling related cases, and divergent rulings have led to inconsistent judicial outcomes.

2. Theoretical foundations for determining copyright infringement liability in generative artificial intelligence

2.1. Technical principles and characteristics of generative artificial intelligence

The three core elements of artificial intelligence generation are algorithms, computational power, and data. Among these, data is the decisive factor in determining whether generative AI constitutes infringement. During data collection and preprocessing, large amounts of data—such as images, texts, and audio—are gathered from various sources. These data serve as the foundation for AI learning, and their diversity, rigor, and scale are crucial to model performance. The main steps include cleaning, labeling, and tokenizing the collected data. The model then predicts outputs based on the inputs, calculates the loss between the predicted results and the true labels, and, after training, learns the probability distribution of the data. In the generation stage, the model samples from the latent space according to this distribution and produces new data through generative processes [2].

In these processes, probabilistic generative models learn the probability distribution of the input data. If the training dataset contains copyrighted works that have been included without authorization, the model will learn their features and patterns. The generated content may thus resemble copyrighted works and infringe upon the reproduction rights of the original authors. If infringing training samples are improperly handled—for example, failing to effectively identify and exclude them—this may indirectly contribute to the generation and dissemination of infringing content.

DOI: 10.54254/2753-7048/2025.LD28196

2.2. Standards for determining copyright infringement liability in generative artificial intelligence

In cases involving copyright infringement by generative AI, the identification of infringing parties is highly complex. The boundaries between developers and users are often blurred: developers provide the technologies and models, while users employ them to generate content, making it difficult at times to determine who should bear responsibility. In the digital environment, AI-generated content can be created and disseminated rapidly, complicating rights enforcement. Furthermore, infringing content may spread widely and displace the market share of original works, causing significant economic loss to copyright holders. In certain cultural industries, for instance, low-cost infringing artworks generated by AI have been noted to disrupt the market for original painters, negatively affecting the creative ecosystem. The large-scale dissemination of infringing content may further disturb cultural market order, dampen creators' enthusiasm, and hinder cultural innovation and development.

According to the Copyright Law, a work refers to intellectual achievements in the fields of literature, art, or science that are original and can be expressed in a certain form. Originality is the key criterion: "independence" requires that the work originate from the author, while "creativity" requires that it embody the author's intellectual creativity, choices, arrangement, and individuality. If a person, without authorization from the copyright holder, engages in activities subject to copyright protection and lacks valid defenses such as fair use or statutory license, infringement is established. Typical infringing acts include unauthorized reproduction, adaptation, or dissemination of others' works. In practice, several key points must be observed. First, determine whether the work falls within the scope of copyright protection; objects not protected include ideas, limited expressions, facts, and works in the public domain. Second, confirm whether the plaintiff is a copyright owner, licensee, or interested party, based on evidence such as copyright registration certificates, manuscripts, or originals. Third, the plaintiff must demonstrate that the defendant had access to the work—for example, where the work was publicly released or where there existed a specific relationship between the parties. A comparison is then made between the protected portions of the plaintiff's work and the corresponding portions of the defendant's work to judge whether substantial similarity exists, both in overall impression and specific content. Finally, possible defenses must be reviewed. Uses of works for purposes such as personal study, research, commentary, or news reporting may, under certain conditions, not constitute infringement. Moreover, in certain legally prescribed circumstances, works may be used without the copyright holder's authorization provided that remuneration is paid, as in the case of reprints by newspapers or broadcasts by radio stations.

3. Current status and issues of determining copyright infringement by generative artificial intelligence in China

3.1. Legislative status of copyright in generative artificial intelligence in China

Internationally, the European Union has adopted a more proactive legislative stance. Although the Artificial Intelligence Act does not directly address copyright infringement, its risk-based regulatory framework offers valuable insights for supervising copyright issues in artificial intelligence. By imposing strict controls on high-risk AI systems, the EU indirectly regulates AI application scenarios that may involve copyright infringement, emphasizing the legality and transparency of data sources. This provides a foundation for subsequent copyright-related legislation.

DOI: 10.54254/2753-7048/2025.LD28196

Domestically, China's existing copyright legal framework is primarily based on the Copyright Law and the Civil Code [3]. The Copyright Law stipulates that works must possess originality and reproducibility, and it explicitly defines infringement acts such as plagiarism and unauthorized use of protected works. However, current provisions remain ambiguous regarding ownership of AI-generated content and fair use of training data. The Interim Measures for the Administration of Generative Artificial Intelligence Services, issued in 2023, is China's first normative document specifically targeting generative AI. It requires service providers to use legally sourced data and prohibits infringement of others' intellectual property rights. Nonetheless, as a departmental regulation, its legal authority is relatively limited, and it lacks detailed infringement criteria, reducing its direct value as judicial guidance.

3.2. Judicial status of copyright in generative artificial intelligence in China

In February 2025, the Hangzhou Internet Court issued the nation's first ruling on an infringement of the right of communication to the public through information networks by a generative AI platform. The plaintiff, the rights holder of the Ultraman character series, argued that the defendant's AI platform allowed users to upload images for training and subsequently generate images substantially similar to Ultraman characters. The court held that although the defendant did not directly generate the infringing content, it provided technical services, established a profit model, and failed to fulfill a reasonable duty of care, thus constituting contributory infringement. In reaching its decision, the court considered factors such as the platform's profit model, degree of technical control, fame of the copyrighted work, foreseeability of infringement proliferation, and subjective fault.

In the "AI-Generated Pattern" case before the Guangzhou Internet Court, the court directly found that AI-generated content constituted reproduction and adaptation of the original work. In this case, the defendant's platform offered AI-based image generation services that produced images similar to copyrighted Ultraman artwork. The court ruled that although the defendant was not the model trainer, its failure to establish keyword filtering or copyright review mechanisms amounted to negligence. Unlike the Hangzhou "Ultraman Case," which emphasized the platform's profit model and duty of care, this case focused on the nature of the generation process, attributing liability based on content generation. This provided an alternative judicial approach for similar disputes.

In domestic judicial practice, similar cases continue to emerge, further enriching and refining the judicial standards for the determination of generative AI copyright infringement. In addition, the Beijing Internet Court, in handling a dispute over copyright of AI-generated images, examined the extent and nature of human involvement in the generation process. In this case, the plaintiff used an AI drawing software, input detailed creative instructions, and repeatedly screened and adjusted the results, ultimately producing a work with a unique artistic style. The defendant used this image in a commercial advertisement without authorization. The court held that although AI played a significant role, the plaintiff's carefully designed inputs and iterative adjustments reflected sufficient individualized intellectual contribution, giving the final image originality distinct from AI presets or similar works. Thus, the plaintiff held copyright over the image, and the defendant's unauthorized use constituted infringement requiring compensation. This ruling emphasized that determining copyrightability of AI-generated content should consider the substantive degree of human participation throughout the creative process, providing a practical benchmark for similar cases.

Another case in Shenzhen focused on the legality of AI training data sources and indirect attribution of liability. A technology company, in developing an intelligent writing assistant, scraped a large amount of unauthorized online literature for model training. Some of the outputs generated by this assistant were found to be highly similar to copyrighted works owned by the plaintiff.

DOI: 10.54254/2753-7048/2025.LD28196

Although the company did not directly reproduce or disseminate the plaintiff's works, the court found that using unauthorized works for training created latent risks of subsequent infringement. Even though AI outputs contained randomness and transformation, the unlawful data source combined with substantial similarity to original works meant that the company could not escape liability. This case underscores that lawful acquisition of training data is the key starting point for avoiding infringement risks, and even where specific infringing outputs are difficult to trace, illegal training data may serve as a crucial basis for determining infringement.

3.3. Existing problems in determining copyright infringement by generative artificial intelligence in China

First, difficulties in defining liability. Article 52 of the Copyright Law clearly stipulates that plagiarism of others' works or unauthorized adaptation, translation, or annotation of works (unless otherwise provided by the law) constitutes copyright infringement and should give rise to civil liability. In the context of generative AI, this provision is critical for determining the legal nature of unauthorized use of copyrighted works in training or content generation.

Second, evidentiary challenges. Traditional copyright law relies on "human intellectual creation" as the core requirement for originality. The "algorithmic generation + human instruction" model of AI challenges this standard, raising doubts about authorship. While the Beijing Internet Court has recognized AI-generated works as copyrightable if they reflect sufficient human intellectual input, it has not clarified how to weigh the relative contributions of "instruction input" and "algorithmic generation." For instance, if users merely input simple keywords, whether the resulting AI-generated content has enough originality to constitute a "work" remains uncertain. In infringement analysis, AI-generated content may avoid direct replication by relying on "style imitation" or "element recombination." For example, if an AI learns Picasso's style and generates new works, the traditional "overall impression" test struggles to determine whether infringement has occurred. In the U.S. case The New York Times v. OpenAI, the plaintiffs argued that AI-generated content diluted the market value of their reporting, but they encountered evidentiary difficulties because they could not demonstrate "verbatim copying."

4. Comparative analysis of copyright infringement determination in generative artificial intelligence

4.1. Case analysis of copyright infringement by generative AI in the United States

As a global frontier of AI technological development, the United States has witnessed a surge of intellectual property infringement lawsuits involving generative AI. On June 23, 2025, U.S. District Judge William Alsup issued a high-profile ruling in which he held that the AI start-up Anthropic's use of copyrighted books for training its models was lawful and protected under the fair use doctrine, thereby exempting the company from compensating authors. This judgment set a precedent in U.S. AI copyright disputes by signaling judicial support for AI companies. Meanwhile, The New York Times filed a lawsuit in the Southern District of New York against OpenAI, alleging that the company had used large volumes of New York Times content without authorization to train its models, thus constituting copyright infringement.

On October 21, 2024, Dow Jones & Company, the parent company of The Wall Street Journal, together with the New York Post, filed suit against Perplexity, a rising AI search firm. In a 42-page complaint, the plaintiffs brought three infringement claims against Perplexity [4]. The lawsuit

DOI: 10.54254/2753-7048/2025.LD28196

primarily alleged that Perplexity had used copyrighted works without authorization in developing its large language model and generating user responses, thereby infringing the plaintiffs' copyrights. The plaintiffs further accused Perplexity of misleading trademark use and trademark dilution. On October 24, Perplexity responded publicly on its website, dismissing the litigation as "short-sighted, unnecessary, and counterproductive," and rebutted several of the plaintiffs' arguments. According to the Database of AI Litigation maintained at George Washington University, this marked the 50th AI-related case filed in the United States.

4.2. Case analysis of copyright infringement by generative AI in the United Kingdom

In the United Kingdom, stock image provider Getty Images filed suit against Stability AI in the High Court of London, alleging that the company had trained its AI art-generation tool using vast numbers of Getty Images without authorization, thereby infringing copyright. On the 16th of the same month, three artists jointly sued Stability AI, Midjourney, and the online art platform DeviantArt, claiming that these entities trained AI models on a dataset of five billion images scraped from the internet "without the consent of the original artists," thus violating the rights of "millions of artists." On January 17, Getty stated that Stability AI had "illegally" extracted millions of images from its website.

When adjudicating such cases, UK courts focus on the legality of the data sources used for AI training. Getty CEO Craig Peters explained: "The reason for this pre-action notice is that Stability AI used others' intellectual property without license or consideration, for its own economic benefit. We do not regard this as fair dealing. The company never entered into discussions with Getty, so we are taking action to protect the intellectual property of both the company and artists." If the acquisition of data violates copyright law—for example, by scraping copyrighted images or texts without authorization—then even if the AI model subsequently performs complex transformations on the data, infringement may still be found. This judicial approach highlights the need to regulate AI data use at the source to safeguard the rights of copyright owners.

However, the traditional fair use principle encounters interpretive difficulties when applied to AI training data, particularly regarding the standard of transformative use, which remains inconsistent. In the U.S. Anthropic case, the judge analogized AI training to the human learning process, reasoning that its purpose was "to create new content rather than to substitute for original works," and therefore applied fair use. Yet in Thomson Reuters v. Ross, the court rejected the transformative characterization of AI training, finding instead that the development of a competing product harmed the market value of the original works. This conflict stems from the fundamental difference between human creativity and AI's "input–output" logic: whereas humans create through independent intellectual expression, AI generates content by fitting patterns across massive datasets, making its "transformative" nature difficult to evaluate under traditional originality standards.

4.3. Comparative analysis of domestic and international approaches to determining copyright infringement by generative AI

In China, although no explicit judicial guidelines on fair use exist, the Hangzhou "Ultraman case" adopted a reasoning that links platform liability to its profit model and duty of care. This partially overlaps with the U.S. approach, where the commercial nature of use affects fair use determinations. However, the absence of unified standards risks inconsistent outcomes in similar cases.

The "training–generation–dissemination" chain of generative AI involves multiple actors, including model developers, data providers, platform operators, and end-users. Current law struggles

DOI: 10.54254/2753-7048/2025.LD28196

to delineate liability boundaries clearly [5]. Technically, developers control the sources of training data but may not directly engage in content generation; platform operators provide generative tools but do not control algorithmic logic; users trigger the generation process through prompts yet cannot fully anticipate the results. This decentralized liability chain makes attribution of infringement difficult. In the Hangzhou Ultraman case, the court held the platform liable for "assisting infringement," but did not address the liability of model developers. In the Guangzhou AI text-to-image case, the court focused on generative conduct at the output stage but avoided the issue of training data legality. By contrast, in the UK Getty v. Stability AI case [6], the court directly identified "data scrapers" as infringers. Chinese judicial practice remains comparatively cautious in apportioning multi-actor liability, making it difficult to implement the intuitive principle of "he who benefits should bear responsibility."

5. Recommendations for improving the attribution of copyright infringement liability in generative artificial intelligence

5.1. Legislative recommendations

First, clarify the rules for determining the ownership of rights in AI-generated content. It is recommended that the revised Implementation Regulations of the Copyright Law include provisions on the copyrightability of AI-generated content. At the same time, the originality standard of AI-generated content should be specified, with the degree of human intellectual input and the novelty of expression as the core criteria. Second, refine the contextual standards for "fair use." For instance, by referencing the U.S. Copyright Act's "four-factor test" and considering the technical characteristics of AI, specific guidelines for "fair use" of training data should be established. Limited use of data in non-commercial scientific research scenarios could be permitted, while wholesale reproduction without authorization, such as in The New York Times v. OpenAI [7], should be restricted. Third, improve the rules for allocating liability among multiple parties. A special chapter on "AI Infringement Liability" could be added to the Civil Code's Tort Liability section, specifying the duties of model developers to verify the legality of training data, platform operators to review generated content, and users to comply with input obligations.

5.2. Judicial recommendations

First, issue guiding cases and uniform adjudication standards. The Supreme People's Court could synthesize typical cases such as the "Ultraman Case" and the "AI-Generated Pattern Case," distilling key factors for determining "fair use" and "substantial similarity." In training data infringement cases, the focus should be on the volume of data used and whether it substitutes for the original work's market; in cases involving generated content, technical evidence such as "algorithm comparison reports" could be introduced to assess similarity and imitation levels. Second, explore proportional liability models. In multi-party infringement cases—such as those involving users uploading infringing data, platforms providing tools, and developers failing to implement filtering mechanisms—liability could be apportioned according to each party's degree of fault and profit [8]. This would prevent any single party from bearing excessive liability or leaving responsibility gaps.

5.3. Technical and industry-level recommendations

First, establish a tiered platform review mechanism. Based on platform influence and technical capacity, differentiated duties of care should be imposed. Large platforms should deploy AI content

DOI: 10.54254/2753-7048/2025.LD28196

filtering systems and keyword blocking in high-risk domains, while smaller platforms may adopt a "manual review + user commitment" model, provided that review records are preserved for future inspection.

Second, promote industry self-regulation and compensation mechanisms. AI enterprises should be encouraged to form self-regulatory alliances and issue Guidelines for Copyright Compliance in Generative AI, setting clear boundaries for data acquisition and content generation. A "copyright compensation fund" mechanism could also be explored, whereby AI enterprises contribute a percentage of revenue to a dedicated fund to compensate creators whose rights are infringed—thus balancing innovation with rights protection.

5.4. International cooperation recommendations

China should actively align with international norms such as the EU's Artificial Intelligence Act and U.S. copyright guidelines. In cross-border infringement cases, the "principle of comity" could be applied: AI products that meet another country's lawful compliance standards may undergo simplified review in China. At the same time, a "Global AI Copyright Dispute Mediation Center" could be established to resolve cross-border disputes through non-litigation channels, reducing compliance costs for enterprises.

Through these measures, copyright holders' legitimate rights can be safeguarded while leaving space for technological innovation, thereby achieving a dynamic balance between "legal regulation and technological development."

6. Conclusion

The rapid evolution of generative artificial intelligence has reshaped the paradigm of content creation while simultaneously challenging the traditional copyright system that is centered on human authorship. With respect to the attribution and allocation of copyright infringement liability, a practical dilemma has emerged: rapid technological development, delayed legal regulation, and exploratory judicial practice. Addressing this requires systematic solutions from theoretical, legislative, judicial, and industry collaborative perspectives. In conclusion, the attribution and allocation of copyright infringement liability in generative AI is not a matter of simply "law adapting to technology" or "technology yielding to law." Instead, it requires a dynamic balance among rights protection, technological innovation, and industrial development. Only by clarifying rules through legislation, unifying standards through the judiciary, strengthening safeguards through technology, and resolving cross-border disputes through international cooperation can we delineate a legitimate and compliant development boundary for generative AI. This will not only safeguard the intellectual achievements of copyright holders but also provide sufficient space for innovation, ultimately fostering the high-quality development of the generative AI industry on a law-based trajectory and injecting lasting momentum into cultural innovation in the digital economy era.

References

[1] Feng, X., & Shen, Y. (2025). Determination of copyright infringement liability of generative artificial intelligence service providers. Rule of Law Studies, 2025(1), 46–58. https: //doi.org/10.16224/j.cnki.cn33-1343/d.20241230.005

[2] Li, Y. (2025). "Risk-based approach" in artificial intelligence governance: From the perspective of the EU Artificial Intelligence Act. Journal of Nanjing University of Posts and Telecommunications (Social Science Edition), 1–9. Advance online publication. https: //doi.org/10.14132/j.cnki.nysk.20250708.001

[3] Wang, L. (2025). Attribution principles and fault determination in generative artificial intelligence infringement. China Law Review, 1–15. Advance online publication. http:

DOI: 10.54254/2753-7048/2025.LD28196

//kns.cnki.net/kcms/detail/10.1210.D.20250702.1616.002.html
[4] Hecan, Z., Yi, C., Guo, P., et al. (2024). Copyright data dilemma of building a high-quality data system for AI: Present situation, coping strategies, and implementation path. Journal of Library & Information Science in Agriculture, 36(9).
[5] Li, T. (2025). Application logic and risk governance of judicial artificial intelligence. Administrative Law Review, 2025(5), 46–62. Advance online publication. http: //kns.cnki.net/kcms/detail/11.3110.D.20250707.1555.008.html
[6] Gaidartzi, A., & Stamatoudi, I. (2025). Authorship and ownership issues raised by AI-generated works: A comparative analysis. Laws.
[7] Vivek, P. (2024). Reimagining copyright: Analyzing intellectual property rights in generative AI. Journal of Information and Knowledge, 327–331. https: //doi.org/10.17821/srels/2024/v61i6/171646
[8] Samuelson, P. (2023). Generative AI meets copyright. Science, 381(6654), 158–161. https: //doi.org/10.1126/science.adi0656