Law and Humanities
ISSN Print: 3079-5117
ISSN Online: 3079-5125
https://doi.org/10.63313/LH.9007

Research on Copyright Infringement Issues of Generative Artificial Intelligence

Lei Jiang
Anhui University of Finance and Economics, Bengbu 233030, China
Email: 1507435247@qq.com

How to cite this paper: Jiang, L. (2025). Research on Copyright Infringement Issues of Generative Artificial Intelligence. Law and Humanities, 1(1), 51–60. ISSN Print: 3079-5117, ISSN Online: 3079-5125.
https://doi.org/10.63313/LH.9007
Published: 2025-05-12

Abstract

This article takes two AI infringement cases involving Ultraman in Hangzhou and Guangzhou as entry points to analyze the infringement disputes of genera-tive artificial intelligence in the stages of data input, model training, and con-tent output. In the Hangzhou case, the platform was found to be liable for con-tributory infringement because users generated infringing content using its tools, and the court recognized fair use during the training stage. In the Guang-zhou case, the platform directly generated infringing content, was determined to be directly in-fringing, and fair use was not applicable. The differences in the judgments of the two places reflect the liability stratification of "technology tool providers" and "content service providers", as well as the balance struck by the judiciary between technological innovation and copyright protection. It is recommended to add spe-cial provisions on fair use of AI training, establish flex-ible judgment criteria such as the "three-step test" and "transformative use", clarify the liability boundaries of different service models, and promote the co-ordinated development of techno-logical innovation and rights protection.

Keywords

Generative Artificial Intelligence; Copyright Infringement; Fair Use; Three-step Test; Transformative Use

Introduction

With the rapid development of generative artificial intelligence technology, AI tools centered around deep learning models have widely penetrated into fields such as content creation, image generation, and text generation. However, the conflict be-tween technological innovation and copyright protection has also intensified ac-cordingly. From the potentially protected works contained in training data, to the substantial similarity between the content output by the model and prior works, and then to the definition of the responsibilities of platform operators for user - generated content, the copyright infringement disputes caused by generative AI throughout the entire chain of data input, model training, and content output are becoming the focus of judicial practice and academic research.

In 2024, two "Ultraman AI Infringement Cases" were successively adjudicated by the courts in Hangzhou and Guangzhou. These cases fully exposed the complexity of generative AI copyright disputes and the differences in judicial judgments. In the Hangzhou case, the platform was found to be liable for contributory infringement because users generated infringing content using its tools, and the court recognized the fair use during the model training stage. In the Guangzhou case, the platform was determined to be directly infringing because it directly generated infringing content, and its fair - use defense was not supported. The responsibility stratification of "technology tool providers" and "content service providers" in the judgments of the two places, as well as the judicial thinking of balancing technological innovation and copyright protection, provide typical examples for clarifying the legal boundaries of generative AI copyright infringement.

This article takes the two "Ultraman AI Infringement Cases" as the starting point, disassembles the focal points of infringement disputes in the three stages of data input, model training, and content output, analyzes the judicial logic of distinguishing between contributory infringement and direct infringement, and explores the application dilemmas of fair - use rules in the AI scenario. The research aims to provide theoretical references for improving the identification standards of generative AI copyright infringement and balancing technological innovation and rights protection through case comparisons and normative analyses, and to contribute to the construction of an intellectual property governance system that adapts to the digital age.

1. Overview of Generative Artificial Intelligence

Generative artificial intelligence is a new type of technology based on deep learning algorithms. It autonomously generates creative new content by analyzing the patterns and laws of massive data. Its core technologies include:

- Neural network architectures: such as Transformer, Generative Adversarial Networks, and Diffusion Models.
- Training paradigms: Adopting the "pre - training + fine - tuning" model. For example, GPT - series models master language rules through large - scale unsupervised learning and then adapt to specific tasks through prompt engineering.
- Generation logic: Based on probability distribution, content output is completed through autoregressive prediction or latent variable sampling.

The training process of generative AI can be refined into multiple links, and each link involves potential legal risks:

- First is data collection and cleaning. That is, publicly available data such as social media graphics and texts, e - books, etc. are retrieved from the Internet, and after pre - processing steps such as deduplication, denoising, and annotation, a training set is formed. For example, the training dataset LAION5B of Stable Diffusion contains 5.85 billion image - text pairs.

- Second is model pre - training. Massive data is used to train the basic model, and hundreds of millions to hundreds of billions of parameters are adjusted through the backpropagation algorithm. This stage requires consuming the huge computing power of thousands of GPU clusters for months of training.
- Then comes domain fine - tuning and optimization. With the help of small - scale domain data such as medical literature and legal documents, technologies like LoRA and Prompt Tuning are used to conduct targeted optimization of the pre - trained model.
- Finally is content generation and deployment. Users input prompt words like "Ultra - clear scene of Ultraman Zero fighting against monsters", and the model generates content by combining training knowledge and delivers it to users through APIs or application interfaces such as Midjourney and ChatGPT.

2. Copyright Infringement Disputes of Generative Artificial Intelligence

2.1. Input End: Infringement Risks in the Data Collection and Training Process

The core disputes at the data input end of generative artificial intelligence mainly focus on the legality of training data and the applicable boundaries of fair - use rules. During the process of data collection and cleaning, if unauthorized copyrighted works are included or protected database resources are illegally crawled, it may trigger multiple infringement risks. AI developers often defend with the claim of "fair use", arguing that data training is for the purpose of technical learning or research. However, in judicial practice, the determination of fair use for commercial training is becoming stricter.

2.2. Output End: Infringement Determination in the Content Generation and Dissemination Process

The copyright disputes at the output end focus on the determination of "substantial similarity" between the generated content and copyrighted works, as well as the definition of the platform's liability for infringement results. When users enter specific prompt words or the platform fails to filter the infringing content uploaded by users, it may trigger infringement risks of the right of reproduction, the right of adaptation, the right of communication through information network, and even the right of portrait. If the generated content directly reproduces the original expression of the original work or forms new content after adapting the original work, it constitutes an infringement of the right of reproduction and the right of adaptation without permission. If such content is publicly disseminated on an open platform, it further infringes the right of communication through information network. In addition, generating virtual images of specific individuals may also involve disputes over the right of portrait.

3. Analysis of Core Issues in Copyright Infringement of Generative Artificial Intelligence

3.1. Whether AI Training and Generated Content Constitute Copyright Infringement

In the Hangzhou case, the first - instance court found that the defendant was liable for contributory infringement of the right of communication through information network. The pictures and cover images generated by users through the platform's LoRA model training were substantially similar to the plaintiff's artistic works. The defendant encouraged users to publish content through a profit - making model but failed to take review measures. It should have known about the users' infringement but did not intervene, which met the elements of contributory infringement. The second - instance court upheld this determination, emphasizing that the defendant failed to fulfill its reasonable duty of care for the infringing content, and the defendant could not prove the originality of the generated content, so the infringement was established.

In the Guangzhou case, the court found that the defendant was directly infringing, violating the plaintiff's right of reproduction, right of adaptation, and right of communication through information network. The website operated by the defendant directly generated pictures that were substantially similar to the plaintiff's works through its AI painting function, digitally reproduced, adapted, and disseminated them without permission. The defendant's defense of "the technology was provided by a third party" was invalid because, as a platform operator, it had a duty of review and profited from "computing power" recharge, showing obvious subjective fault.

The difference in the types of infringement determined in the two cases stems from the differences in AI service models. In the Hangzhou case, the platform provided training tools for users, and the infringing content was generated by users, so it was determined as contributory infringement. In the Guangzhou case, the platform directly generated infringing content, so it was determined as direct infringement. This reflects the liability stratification of "technology tool providers" and "content service providers" in the judiciary. The former needs to prove that they "knew or should have known about the infringement but failed to stop it", while the latter, due to direct infringement, needs to bear strict liability.

3.2. Whether the Use of Others' Works in the AI Training Stage Constitutes Fair Use

In the Hangzhou case, the first - instance court held that the use of others' works in the training stage of generative AI constituted fair use. Its training purpose was to learn and analyze the style of works for transformative creation. For example, when users trained the LoRA model with Ultraman pictures, they only extracted features without displaying, reproducing, or disseminating the original works, which was in line with the legislative spirit of the Copyright Law and did not harm the rights of

the right - holders. The second - instance court supported this determination, emphasizing that "non - commercial learning purposes" and "not publicly disseminating the original works" were the key, and the "transformative nature" of the use behavior was important. The training service of this platform was in line with the direction of technological innovation and did not constitute infringement.

In the Guangzhou case, the court did not recognize fair use and directly determined infringement based on "substantial similarity". The defendant used the plaintiff's works for training and generated similar pictures without permission and also profited through a paid function, which did not meet the "non - commercial" and "not harming the rights of the right - holders" requirements of fair use, so the fair - use defense could not be applied.

The first and second instance judgments of the Hangzhou case clearly defined the boundaries of fair use in the AI training stage in judicial practice, which is of great significance. Its core lies in "learning - based use" and "non - publicity". In the Guangzhou case, because the platform directly generated infringing content and made a profit, the fair - use rules did not apply. In the future, the judiciary should further refine the standards. For example, distinguish between "basic model training" and "specific IP - targeted training". The former may be more likely to constitute fair use due to diverse data, while the latter, if it is for a specific work, may be determined as infringement.

3.3. How to Define the Liability Boundaries of Platform Providers for Users' Infringing Acts

In the first - instance judgment of the Hangzhou case, the court determined that the defendant should bear the liability for contributory infringement, and the liability was only for the infringing content at the "output end", such as deleting the infringing pictures and LoRA models published by users and stopping related services. For the users' unpublished training data, considering that the training stage may constitute fair use and the platform's obligation to ensure data security, it was not required to delete all materials, but only to prevent the public dissemination of infringing content. The second - instance court upheld the original judgment, emphasizing that the platform had a higher duty of review for "obviously perceivable infringing content". The defendant needed to bear the liability because it failed to fulfill this obligation, and the scope of liability followed the principle of "minimal necessity".

In the Guangzhou case, the court found that the defendant should bear direct liability for infringement, covering "training data and generation functions". The court required the defendant to delete the relevant materials in the training dataset and block the website from generating similar pictures because the defendant directly used the plaintiff's works for training and provided infringing generation services. There was no room for a fair - use defense, and it was necessary to stop the infringement from the source and during the process.

The two cases reflect the differences in the liability definition of the judiciary for different AI service models. The Hangzhou case, for the "user - generated content" platform, only held the platform accountable for the publicly disseminated infringing results, reflecting a restrained review and focusing on "post - event governance". The Guangzhou case, for the "platform - generated content" platform, imposed strict liability, requiring the complete cut - off of the infringement chain and focusing on "pre - event prevention". This distinction is in line with the judicial concept of "balancing technological innovation and rights protection".

3.4. Whether the Platform Business Model Constitutes Unfair Competition

In the first - instance judgment of the Hangzhou case, the court believed that the defendant's LoRA model training service was a new type of generative AI service, which provided users with personalized creation tools and was technologically neutral. The infringement by some users fell within the scope of adjustment by the Copyright Law, and the Anti - Unfair Competition Law did not duplicate the evaluation. Moreover, the defendant did not actively induce infringement, its business model was in line with industry practices, and it did not violate the principle of good faith or disrupt the market order, so it did not constitute unfair competition. The second - instance court upheld the original judgment and pointed out that the prerequisite for the application of the "principle of technological neutrality" was that the platform service had "substantially non - infringing commercial uses". In this case, the defendant's platform was mainly used for users' independent creation, and a small number of infringement cases did not affect the legitimacy of its business model. The relevant disputes were essentially "substantial similarity" issues in the Copyright Law and did not require regulation by the Anti - Unfair Competition Law.

The first and second instance judgments of the Hangzhou case clarified the rules for determining "technological neutrality" in unfair competition. When the platform business model is innovative and generally applicable, the infringement caused only by users' behavior does not necessarily constitute unfair competition. However, it is necessary to be vigilant against "technology abuse", such as the situation of profiting from popular IPs. Since the defendant in this case did not engage in such behavior, it reflected the judiciary's tolerance for technological innovation.

4. Suggestions for Improving the Copyright Infringement Issues of Generative Artificial Intelligence

4.1. Adding Special Provisions on Fair Use of AI Training

In the technical chain of generative artificial intelligence, data training is a core link. However, due to the lack of clear fair - use rules for AI scenarios in the current Copyright Law, there are differences in judicial judgments. Therefore, it is necessary to add targeted provisions to the law to clarify the boundaries and conditions for fair

use of AI training. First, the legitimacy of "learning - based use" should be recognized. When an AI model is only used to analyze abstract elements such as the style, structure, and characteristics of a work, and does not publicly disseminate the original work or substitute its market value, it can be considered fair use. This "non - replicative learning" is a necessary path for technological innovation. A total ban would seriously hinder the development of AI.

Secondly, dual standards of "purpose limitation" and "effect evaluation" should be set for fair use. Non - commercial training should be given greater tolerance, while commercial applications need to meet the requirements of "transformative use", that is, the generated content must create new forms of expression or functional value, rather than directly reproducing the core expression of the original work. If a platform conducts in - depth imitation training for a single IP and generates substantially similar content, since it directly damages the licensing revenue of the right - holder, fair use should be excluded, and prior authorization is required. By clarifying these boundaries through legislation, it can not only prevent "technology abuse" but also leave room for AI to learn from massive legal data.

4.2. Establishing a Unified and Flexible Legal Application Standard

The determination of copyright infringement in generative artificial intelligence needs to rely on the "three - step test" of the Copyright Law and integrate the foreign "transformative use" theory to construct a legal application system that combines rigidity and flexibility. The "three - step test" is dynamically applicable to AI technology scenarios: in the first step of "specific circumstances" limitation, if AI training and content generation are for the purpose of "technology development", they can be included in the scope of fair - use examination, while commercial utilization for the purpose of directly replicating the original work is excluded. In the second step of "not affecting normal use", the market substitution effect needs to be considered. Non - commercial learning or generating non - core expression content is presumed to cause no harm, while behaviors that directly substitute the commercial use of the original work are determined to affect the interests of the right - holder. In the third step of "not harming the interests of the right - holder", subjective fault needs to be considered. For example, if a platform induces infringement through a paid model, even if the technology is neutral, fair use may be excluded due to fault.

The "transformative use" theory, as a supplement, strengthens the evaluation of the innovativeness of the generated content, requiring that new expressions should endow the original work with new functions or social significance, rather than simple imitation. In judicial practice, "transformativeness" can be used as a reference for judging "whether it affects normal use". Significantly innovative content can be leniently recognized even if it involves commercial use, leaving room for technological innovation. Conversely, if it is only a minor modification or pixel - level replication of the original work, the theory is not applicable due to the lack of new value.

The determination of platform liability is stratified according to the "service model": technology tool providers only bear the liability for contributory infringement for "publicly infringing content of users that they know or should know about", reflecting the limited duty of care under "technological neutrality". Content service providers, because they actively generate content or control training data, need to undergo dual review of the "three - step test + transformative use". If they cannot prove compliance, they are presumed to be directly infringing and bear liability. This stratified determination not only adheres to the legal bottom line but also adapts to technological diversity through flexible standards, avoiding the suppression of innovation vitality.

4.3. Promoting Industry Self-regulation and International Collaboration

Industry self - regulation is an important supplement to legal regulation and can reduce infringement risks from the source. On the one hand, it is necessary to promote the establishment of a collective copyright authorization mechanism between rights - holder organizations and AI enterprises to solve the problem of "massive authorization" through standardized license agreements. For example, an "AI compliant training data pool" can be established to integrate public - domain works, open content voluntarily authorized by rights - holders, and authorized materials determined through collective negotiation. Enterprises can pay license fees according to the number of model calls or revenue ratios, which not only guarantees the income of rights - holders but also reduces the compliance costs of enterprises. On the other hand, platforms should be guided to deploy technical filtering measures, such as developing an "infringement prompt word library" to automatically intercept or provide risk warnings for high - risk instructions like "generate Ultraman" and "replicate a certain painter's brushstrokes", and adding digital watermarks to generated content to record metadata such as the model source and the authorization status of training data, facilitating rights - holders to trace the source of infringement. These measures can not only enhance users' copyright awareness but also provide technical evidence - taking support for subsequent disputes.

At the international level, it is necessary to actively draw on foreign experience and participate in rule - making. The "transformative use" theory in U.S. judicial practice and the data transparency requirements of the EU's Artificial Intelligence Act provide useful references for China. China can promote the establishment of an international consensus on "fair use of artificial intelligence training" under the framework of the World Intellectual Property Organization (WIPO), clarify the boundaries between "idea borrowing" and "expression replication" in technology development, and avoid governance loopholes caused by differences in rules among countries. For cross - border IP infringement disputes, law enforcement cooperation memorandums should be signed with major technology - exporting countries to share lists of infringing platforms and technical testing standards, and explore a "cross - border

copyright fast - arbitration mechanism" to solve problems such as difficult cross - border evidence collection and jurisdiction conflicts. Through the combined efforts of industry self - regulation and international collaboration, a healthy domestic innovation ecosystem can be created, and China can build rule - making power for its generative AI industry to participate in global competition.

Conclusion

The copyright infringement disputes caused by generative artificial intelligence essentially represent a profound game between "innovation freedom" and "rights definition" during the process of digital technology reconstructing the creative ecosystem. From the disputes over the legality of training at the data input end to the stratification of infringement liability at the output end, the differences in the judgments of the two "Ultraman AI Infringement Cases" reveal the challenges of adapting technological characteristics to legal rules. It is necessary to avoid suppressing technological innovation due to over - regulation while also preventing "technological neutrality" from being misused as a tool for infringement. Through differentiating the liability boundaries of "technology tool providers" and "content service providers", judicial practice has initially established a governance framework that combines "process control" and "result - based liability", providing a reference for judging "substantial similarity determination" and "fair - use defense" in similar cases.

At present, the governance of generative AI copyrights in China faces the dual challenges of lagging rule - making and rapid technological iteration. The suggestions put forward in this article, such as adding special provisions for fair use, establishing flexible legal application standards, strengthening industry self - regulation, and promoting international cooperation, aim to achieve a dynamic balance between innovation incentives and rights protection through a multi - dimensional approach of "defining boundaries with laws, empowering with technology, and building an ecosystem through collaboration". However, with the continuous emergence of frontier issues such as the "recognition of the work attributes of AI - generated content", "rules for rights attribution", and "jurisdiction over cross - border infringement", future research needs to further focus on the in - depth interaction between technological innovation and the copyright system, and explore new paradigms for intellectual property governance in the era of "human - machine collaborative creation".

References

[1] Chen,J.&Zhang,W.R.(2024).Reconsideration of the fair use system for genera-tive artificial intelligence[J].Journal of Changchun University,(5).

[2] Chandrakar,V.(2024).'From Canvas to Code:Analysing Generative AI and Fair Use Through the Lens of the Andy Warhol Verdict'[J].Journal of Intellectual Property Studies,8(1).

[3] Myers,G.(2023).'Artificial Intelligence and Transformative Use After Warhol'[J].Washington&Lee Law Review Online,81(1).

[4] Murray,M.D.(2023).'Generative AI Art:Copyright Infringement and Fair Use'[J].SMU Science and Technology Law Review,26(2).

[5] Wan,Y.(2021).Dilemmas and solutions for the fair use system in copyright law in the era of artificial intelligence[J].Social Science Journal,(5).

[6] Zhang,M.(2023).An interpretive study on the application of copyright fair use in text and data mining[J].Publishing Research,(2).