

Research on the Subject of Copyright Infringement Liability for Generative Artificial Intelligence

Shenggang Wang

China Jiliang University, Hangzhou, Zhejiang, China

Abstract: The explosive development of generative artificial intelligence technology, while bringing about a revolution in content creation, is also profoundly challenging the boundaries and balance of the current copyright legal system. In the algorithmic black box environment, the diverse participation of generative artificial intelligence developers, service providers and users forms a responsibility gap, leading to the risk of diversified responsibility subjects and blurred responsibility boundaries. This study focuses on the latent copyright infringement risks throughout the entire chain of generative artificial intelligence applications. Based on risk identification and cause analysis, by deconstructing infringement disputes, institutional dilemmas, and judicial practices, it reveals the structural failures of current rules in application and proposes a hierarchical governance path, with the aim of achieving a dynamic balance between promoting creators' rights and interests and promoting artificial intelligence innovation.

Keywords: Generative Artificial Intelligence Copyright Infringement; Algorithm Black Box; Responsible Entity

1. Introduction

With the in-depth application of generative artificial intelligence technologies represented by ChatGPT, OpenAI and DeepSeeK in various fields, human society is undergoing a paradigm shift from "content consumption" to "content generation". However, the contradiction between technological progress and the lag of laws has erupted intensively in the field of copyright. Although the "Interim Measures for the Administration of Generative Artificial Intelligence Services" has been officially implemented, a comprehensive review of the entire "Measures" shows that the issue of copyright infringement related to generative

artificial intelligence remains unanswered. The research on this issue concerns the allocation of legal responsibilities and the design of relevant systems in the context of artificial intelligence, and holds significant practical and theoretical significance. Therefore, it is necessary to sort out the governance predicament of copyright infringement in AI-generated content and clarify the possible copyright infringement behaviors that may exist during the generation process.

2. Analysis of Infringement Risks Involving Multiple Subjects

2.1 Developer: The Training Work Involves Direct Infringement

The risk of copyright infringement liability for developers mainly stems from the works used in the training of AIGC models. The sources of database training can be summarized into the following three aspects: First, content from the public domain; The second type is content obtained through one-on-one authorization agreements with copyright holders or through legal authorization by collective management organizations. The third type is the network data crawled using "web crawler" technology. If the work input into the model for pre-training is obtained without the permission of the author of the training data source work, there is a possibility of copyright infringement. The reason is that in the current fair use system, in copyright legislation and judicial practice, if enterprises use others' published works for scientific research purposes, they can defend against infringement through the fair use system. However, for generative AI developers, it is difficult to effectively separate and analyze the commercial and scientific purposes of training data, and the two coexist [1]. Therefore, when developers obtain training data through methods such as crawling without the author's permission and without paying remuneration, there is a possibility of direct copyright infringement. When developers acquire training data, if they

acquire training data sources illegally obtained by other entities, there is a possibility of indirect copyright infringement.

2.2 Service Provider: The Main Type of Infringement is Due to Violation of the Duty of Care

According to the stage at which infringement occurs, it can be divided into the training stage, the generation stage and the removal stage. Service providers who use protected works without permission during the model training process may bear tort liability. The reasons for such infringement are quite different from those of the previous developer and will not be elaborated separately. The situations where service providers are most likely to be held accountable for infringement mainly come from the generation stage and the removal stage. The generation stage is mainly characterized by the content output by the generative AI being substantially similar to the prior work protected by copyright. At present, judicial practice defines the nature tendency of generative artificial intelligence service providers as content service providers, and the "technology neutrality" principle cannot be applied. Therefore, service providers may bear either direct tort liability or indirect tort liability. It is worth noting that socket service providers that directly invoke third-party models without making compliance modifications may bear joint and several liability for third-party infringement. This is because there is no substantial difference from the traditional provision of network services. Network service providers may either directly infringe upon the infringing content on their websites or, due to "knowing or should have known", constitute contributory infringement[2].

2.3 User: Mainly Induced Infringement

There is a legal risk for users to conduct data training on AIGC. When the data involves others' works, if the user uses the data for personal learning and research purposes to help improve their own learning and research efficiency, it can be regarded as fair use and there is no copyright risk. If a user uses another person's work as data for economic benefits, it clearly constitutes the nature of commercial purposes under copyright law, and it should be determined that the user has infringed upon the copyright of the original work's author. In

addition, during the AIGC training data input stage, there is another information collection situation where users input data to ask questions or give instructions, and the AIGC collects the data for training purposes. If the user is aware that AIGC uses the data provided by the customer for processing and computing, it can be determined that the user has assisted in infringement and shall bear the corresponding liability for infringement. Conversely, even if the user directly provides data to others, the user does not need to bear any responsibility for the infringement of the stored data used by the AIGC developer.

3. Theoretical Disputes and Sources Regarding the Subjects of Tort Liability

3.1 The Responsible Entity for Artificial Intelligence Said

The theory of the fully responsible subject of artificial intelligence holds that whether it is regarded as a legal subject is based on whether it is endowed with the identity of personality or the capacity for rights, rather than whether the responsible subject has consciousness or will. Moreover, the discussion on whether to grant artificial intelligence the status of a legal subject should be free from the shackles of anthropocentrism. Some scholars believe that granting artificial intelligence the status of a responsible subject can solve the current problem of difficult fair distribution of copyright infringement liability. When copyright infringement occurs due to the reasons of generative artificial intelligence itself, at this time, the developers, service providers and users of the generative artificial intelligence are all at fault. If developers, service providers and users are involved in bearing the liability for copyright infringement, it will lead to unfair distribution of responsibility. At this time, if artificial intelligence itself can bear the liability for infringement, this problem can be ingeniously solved. The theory of limited liability subjects for artificial intelligence holds that after entering the GPT-4.0 era, generative artificial intelligence has already become a strong artificial intelligence entity, possessing both volitional and behavioral capabilities, and is capable of having a certain simulated personality, bearing limited legal liability. Additionally, it can grant certain interest rights to generative artificial intelligence, enabling the

artificial intelligence itself to have certain benefits. When artificial intelligence itself causes copyright infringement, the artificial intelligence itself may bear the corresponding limited liability, and other entities shall bear the corresponding infringement liability through the principle of "piercing the veil of artificial intelligence". Admittedly, domestic scholars are mainly influenced by the theory of artificial intelligence tools and currently mostly hold a negative attitude towards the above two viewpoints.

3.2 Generative Artificial Intelligence Service Providers Say

Whether it is ordinary robots, animals, or the first generation of weak artificial intelligence, they are all regarded as tools, agent software and other objects[3]. This view holds that it is unrealistic for artificial intelligence to possess independent property and responsibility capabilities. Artificial intelligence is controlled by humans and is an extension of human capabilities. "It has neither endogenous behavioral capabilities nor endogenous rights capabilities[4]." The popular product liability theory in the academic circle also shares similarities with the service provider theory. It points the responsibility subject to the producer of generative artificial intelligence. This theory regards generative artificial intelligence as a product. When the content generated by artificial intelligence causes damage to the copyright of prior works, the producer of generative artificial intelligence should be the responsible subject. However, although product liability and the responsibility of network information producers are different, the responsibility subjects both point to service providers, which are borne by the service providers of generative artificial intelligence. It is undeniable that this has certain positive effects. In order to avoid bearing strict product liability, the service providers of artificial intelligence strive to reduce the risk of copyright infringement during the generation process of generative artificial intelligence. To guide its products towards legality and compliance to the greatest extent possible, at the same time, as the party that gains benefits from artificial intelligence, it is reasonable for the provider of generative artificial intelligence services to bear liability for infringement. However, it is still unreasonable to have the providers of generative

artificial intelligence services as the responsible parties. The liability subjects for copyright infringement of generative artificial intelligence are far more complex, including artificial intelligence developers, providers, and users, etc. If only the providers of generative artificial intelligence services bear the liability for copyright infringement, it would overly burden the service providers. This will be detrimental to the development and growth of artificial intelligence generation technologies and products in our country. Obviously, both theories have certain merit. The two viewpoints can be said to be evenly matched. Even in the context of generative artificial intelligence, they can be applied. It is precisely for this reason that the subject of infringement liability for generative artificial intelligence remains undetermined.

3.3 Source of Dispute: Legislative Deficiencies Have Led to the Breakage of the Chain of Responsibility

The entire value chain of the artificial intelligence generation industry mainly includes three subjects: generative AI developers, service providers, and users. Each subject corresponds to different links in generative AI and has different risks of copyright infringement. However, At present, the "Interim Measures" that have been promulgated in our country only focus the responsibility subject on the provider of generative artificial intelligence services. In addition, The "Basic Requirements for Security of Generative Artificial Intelligence Services" only applies the requirements for corpus security, model security and security measures to GAI service providers. This document was jointly drafted by multiple influential institutions in the fields of research and development, application, and transformation of generative artificial intelligence services, including the China Electronics Standardization Institute, the National Computer Network Emergency Response Technical Team/Coordination Center, Beijing Zhongguancun Laboratory, Zhejiang University, and Baidu. It is the first domestic specification specifically targeting the security field of generative AI. It has a certain degree of authority and can be regarded as a consensus reached by the industry in this regard. Therefore, the current governance model for the copyright infringement liability subjects of generative artificial intelligence shows the characteristics

of partial governance. This governance model that pursues the liability of local subjects for infringement is difficult to cover the possible copyright infringement risks involved in the behavior of generative artificial intelligence, and the division of liability among different subjects is also not fair enough.

4. Limitations of Judicial Trial Practice

Although there are no shortage of theories about the responsibility subjects of artificial intelligence, in current judicial practice, "anthropocentrism" is adhered to. Generative artificial intelligence models do not possess free will and are not legal subjects; they are merely creative tools. The field of generative artificial intelligence mainly includes three subjects: developers, service providers, and users. Each subject corresponds to different links in generative artificial intelligence and has different risks of copyright infringement. However, in the limited judicial trial practice and related precedents, such as the Hangzhou Ultraman case, the Guangzhou Ultraman case, the "Qingyu Nian" case, the copyright infringement dispute between Song Moumou and a certain technology company in Nanjing, and the "AI Companion" software infringement of personal rights case, the service provider is the responsible subject, and the behaviors of developers and users have not been deeply determined, especially when the infringement is triggered by the user's upload. How are the responsibilities of service providers and users defined? The possible reason is that compared with the situation where large-scale users bear tort liability in a scattered manner, service providers can control and handle infringing content in a centralized way, and the cost of preventing and resolving the risk of infringement is lower. It is in line with the principle of economic efficiency for service providers to assume responsibility for the consequences of infringement as "lowest-cost evaders"[5]. Therefore, at the current stage, the centralization of the responsible entities will lead to the separation of rights and obligations as well as the absence of developers.

5. Hierarchical Accountability Centered on "Hazard Control"

5.1 Source Control by Developers

Modern society requires the conduct of certain

dangerous activities, which is particularly important for technological progress. Therefore, the existence of danger is permitted by law. As for the occurrence of danger, if the victim has difficulty obtaining compensation opportunities, it is truly an unfair phenomenon in society. Therefore, to address such unfair issues, the unfortunate consequences of damage should be reasonably distributed. There are mainly four theoretical reasons for this: The first is that danger has been initiated; The second is risk control; The third is the reward for danger; The fourth is the dispersion of dangers[6]. According to the theory of hazard control power, the actor is responsible for controlling the danger that he can control and dominate[7]. Generally speaking, the closer a subject is to the source of danger, the stronger its ability to control infringement risks, and the lower the unit cost of preventing infringement is usually. Theoretically, the duty of care it should bear should be higher[8]. In the entire life cycle of generative artificial intelligence, the human factor is mainly reflected in the following aspects: First, humans design generative artificial intelligence algorithms; second, humans summarize and process data and feed it to the generative artificial intelligence; third, humans instruct the generative artificial intelligence to output content; fourth, humans disseminate the generated content[9]. Generative artificial intelligence, from model development to application, may involve copyright infringement. The involved actors are diverse, and the actions taken by each actor are closely related to the final AI-generated content. Different liability subjects should be clearly defined based on different infringement scenarios. The developers of artificial intelligence are the cornerstone of the development and construction of artificial intelligence platforms, the main body for screening and controlling the data sources of artificial intelligence, the primary node in the governance structure, and the controller closest to the hazard source. The generative artificial intelligence they develop has powerful computing power, which gives it an inherent advantage in identifying and avoiding copyright infringement. It is precisely for this reason that The copyright infringement by developers mainly occurs in the control of algorithms. Developers should optimize algorithms within their technical level to avoid excessive learning of a single prior work, which may lead to

copyright infringement.

5.2 Reasonable Duty of Care of the Provider

GAI service providers act as a platform connecting developers and users. As a bridging entity, they decide whether to introduce artificial intelligence to the market. After entering the market, they continuously influence the use of products and can take preventive measures to avoid damage. However, artificial intelligence has a certain degree of autonomy. The main problem brought about by autonomous capabilities is the decline in human control over intelligent machines. The behavior of intelligent machines has a certain degree of unpredictability, making it impossible to predict how they will operate or the results of interaction and learning with third parties[10]. However, control and domination are one of the logical starting points of the theory of legal liability. "The basis for human beings to be held responsible for their actions lies in having a certain degree of control over such actions[11]." Therefore, GAI service providers mainly bear indirect liability for infringement, which is manifested as follows: At the input end, GAI service providers may cause indirect infringement due to their failure to fulfill a reasonable duty of care over the source of training data; at the output end, GAI service providers fail to promptly detect and handle obviously infringing GAI, while such content has already been disseminated and used. Of course, it cannot be ruled out that it actively participated in or carried out direct infringement acts that infringed upon others' Copyrights.

5.3 Minimum Control Obligations of the User

As the final link in the content output stage, the input instructions of the generative artificial intelligence user and the generation capability of the GAI service jointly act on the infringing content. Therefore, the guiding role of users in the selection and direction of input content during the generation of GAI cannot be ignored. Thus, the issue of users' weights in the division of responsibilities should be explored. There are two modes for users of generative artificial intelligence in the content generation stage: command-based generation and processing-based generation. If the users of artificial intelligence generate content through directive methods, then in the content generation stage, the only existing behavior is the

computational behavior of the algorithm, and the infringement results are also generated by the algorithm. In the case where the contribution degree cannot be distinguished and the GAI service users do not subjectively induce the generation of infringing content, the main force of the infringing content comes from the generation ability of the GAI technology itself. At this point, the infringement liability of artificial intelligence developers and the indirect liability of GAI service providers should be pursued. If the processing generation method is adopted, then in the content generation stage, there is not only the behavior of algorithmic calculation, but also the behavior of people inputting works into the program. For the investigation of infringement of this behavior, attention should be paid to the behavior of users inputting works. In principle, with the current mainstream technical solutions, if images are obtained only through simple input, then this kind of generation is basically single and one-off, showing only highly random content to specific users each time. Therefore, there is no requirement for the information network dissemination right that "the public can obtain the work at the time and place they choose", and it is even more impossible for there to be a "stable" source of infringing work dissemination for a specific right work. However, if another person's work is used without authorization and does not comply with the fair use and legal licensing system, and has already been disseminated and used, the liability for infringement of such behavior shall be borne by the user of artificial intelligence.

6. Conclusion

With the continuous iteration of generative artificial intelligence technology, the control of service providers over the generated content is getting weaker and weaker. Especially under the new trend that various general models are becoming the foundation of new infrastructure, the two main bodies of artificial intelligence developers and service providers are also showing a transformation trend from the current "integration" to "independence". This means that even though the service provider bears the liability for copyright infringement, it does not necessarily mean that it has the ability to adjust and correct the underlying algorithm. Secondly, copyright infringement that may exist in different links is not only related to local

subjects. It is unfair to only take service providers as the responsible subjects. Unlike traditional artificial intelligence, in its design and operation process, the subjects that may bear the liability for copyright infringement present diversified, dynamic and contextualized characteristics. Different subjects will bear different responsibilities in different scenarios, making it difficult to simply define the boundaries of the responsible subjects. The copyright infringement risk of GAI does not only come from service providers. It may also originate from AI developers and end users. For instance, when a user collects another person's work as question information by GAI and uses it as training data, the definition of copyright liability between the user, the developer, and the service provider should not only take into account the user's original data provision behavior but also the assistance responsibility and substitution responsibility of the developer and the service provider.

References

- [1] Song Weifeng. Copyright Risks and Regulatory Reshaping of Generative AI Training Data Sources: A Case Study of ChatGPT Application Scenarios. *Journal of Dalian University of Technology (Social Sciences Edition)*, 2020, 46(02): 100-110.
- [2] Xu Wei. Determination of Tort Fault of Generative Artificial Intelligence Service Providers. *Law*, 2024, (07): 110-124.
- [3] Beck S. The problem of ascribing legal responsibility in the case of robotics. *AI & society*, 2016, 31: 473-48 1.
- [4] Hao Tiechuan: "Do not Imagine or Overestimate the Impact of Artificial Intelligence on the Rule of Law", *Legal Daily*, January 3, 2018.
- [5] See Guido Calabresi, *Concerning Cause and the Law of Torts: An Essay for Harry Kalven, Jr.*, 43 U. Chi. L. Rev. 69 (1975-1976).
- [6] Wang Zejian. *Infringement Acts*. Beijing: Peking University Press, 2009, p. 521.
- [7] Feng Xiaoqing, Shen Yun. Determination of Copyright Infringement Liability of Generative Artificial Intelligence Service Providers. *Rule of Law Studies*, 2025, (01): 46-58.
- [8] Si Xiao. Establishment of Intellectual Property Care Obligations by Network Service Providers. *Legal Science (Journal of Northwest University of Political Science and Law)*, 2018(1):78-88.
- [9] Wu Changhai, Huang Jingyi. Determination of Copyright Infringement by GAI Service Providers from the Perspective of Interest Balance. *Intellectual Property*, 2025, (03): 88-104.
- [10] Weston Kowert, "The Foreseeability of Human-Artificial Intelligence Interactions," *Texas Law Review*, Vol.96, No.1, 2017, pp.183, 192.
- [11] John Martin Fischer and Mark Ravizza, *Responsibility and Control*, Cambridge: Cambridge University Press, 2000, p. 240.