

Research on the Copyright Recognition of Artificial Intelligence Generated Content

Zihao Fang *

The Literature College, Heilongjiang University, Harbin, China

* Corresponding Author Email: 20206515@s.hljy.edu.cn

Abstract. Generative artificial intelligence undergoes statutory copyright licensing of pre-existing data and the training of models to input commands, execute algorithms, and output content. The constant involvement of human intelligence and the creative model of human-computer collaboration determine that artificial intelligence cannot be the subject of copyright. This gives rise to three categories of subjects: Artificial intelligence users, developers and investors. Artificial intelligence always involves human intelligence in the content creation process, and artificial intelligence-generated content has an original contribution as well as a minimum level of creativity from users, formally constituting works under the copyright law. The rules of copyright attribution for artificial intelligence-generated works should be prioritized according to the contractual agreement, If there is no contractual agreement, the copyright protection mode of the copyright law should only extended to AI users. The object of copyright protection should be limited to the artificial intelligence-generated works themselves, and the strength of copyright protection should be in line with the user's originality of the contribution to the work.

Keywords: AI-Generated Content; Copyright; Works; Originality; Intellectual Achievements.

1. Introduction

At the end of 2022, the U.S. company OpenAI launched ChatGPT, a generative artificial intelligence (GAI), marking a new stage in the development of AI technology. Compared with early artificial intelligence (AI), ChatGPT is highly autonomous and independent. The operation of its algorithms and the output of its content no longer depends on real-time human control. It can autonomously generate complex content in a short period of time with only simple instructions. A series of issues, such as how AI generation should be distinguished from human creation in comparison with creation in the traditional sense, whether AI-generated content (AIGC) is in line with the provisions of copyright law for works compared to works in the traditional sense, and what legal effects will be produced if AIGC is recognized as works, have attracted the attention of both the theoretical and practical communities. In response to the challenges to the copyright law system posed by the development of artificial intelligence, a study on the recognition of AIGC in relation to copyright will help resolve the controversial issues mentioned above and provide a basis for studying the attribution of rights to AIGC, the content of the rights and obligations, the limitation of the rights, and the liability for infringement.

AI autonomously generates content is merely a fact, but the legal relationships involved are quite complex and controversial. Therefore, this paper examines the issue from the following ideas. Firstly, the legal nature of the process of AI learning prior data is a prerequisite for discussing the copyrightability of AIGC. Secondly, based on the theory of civil subject, there is an insurmountable gap between the creation of human intelligence and the autonomous generation of AI. Again, based on the first two points, combined with the constitutive requirements of the work, AIGC can be clearly classified and effectively defined. Finally, the object of the copyright points to the attribution of rights, filling in the content of the copyright legal relationship regarding AIGC.

2. The Legal Nature of AI-Learning from Prior Data

When it comes to how AI functions, the training of AI models involves the input and deep learning of pre-existing data. Analyzing the copyrightability of AIGC certainly entails considering two sides of the same coin: the object of copyright and the attribution of copyrights, which are essentially two aspects of the same issue. If AIGC is directly analyzed solely from the perspective of objectivity of originality, it actually overlooks the impact of AI's working principles on the copyrightability of AIGC.

The training of GAI models begins with the utilization of prior data. Under PRC's current legal system, AI's use of pre-existing data involves the issues of fair use, statutory licence use, and the tort liability. In the case of training data, the adoption of an economically uncompensated fair use system is likely to result in frustrating the incentives of the data source, thereby upsetting the balance between motivating creation and facilitating dissemination, and ultimately shaking the foundations of fair use based on the incentive theory. Therefore, AI should adopt a statutory licensing system for training data to compensate the data source and maintain a balance between incentives and creativity [1].

Once the statutory license for AI training data has been established, it is crucial to explore the connection between the content of the training data and copyrightability. The training data used for AI can fall into various categories: it may be in the public domain, still within the statute of limitations for copyright protection, or subject to prohibition by law or regulation. This perspective focuses on the protection and limitation of training data. Training data that falls within the statute of limitations for copyright protection possesses both a personal right and a property rights, highlighting the importance of understanding the content of the rights associated with the training data. Considering these aspects, it is essential to clarify the statutory licensing system for GAI concerning the use of training data. Works protected by the copyright law and already published can be subject to the statutory licensing system, excluding those containing trade secrets. The scope of the statutory licence is linked to the property rights of the data course's copyright. In cases involving personal rights related to copyright, considerations should be based on individual interests and adhere to the principles of limited purpose and minimum scope [2].

Analyzing the data-learning principle of AI and its legal nature is of great significance in proving the copyrightability and originality of AIGC. On the one hand, it provides support for the necessary "selection and arrangement" advocated by the subjectivism of originality based on data sources and machine learning principles. On the other hand, it establishes the prerequisites for the objectivism of originality, which analyses the minimum originality of AIGC compared with the pre-existing data and AIGC meets all the formal requirements for constituting a work.

3. AIGC and the Subject of Copyright

The discussion of the copyrightability of AIGC should not only start from the process of autonomous generation by AIGC but also need to focus on the relationship between AIGC creation and creation by natural persons, and to analyze what kind of position AIGC holds in the process of creation, and whether it has the status of a creative subject. If it does not have subject qualification, then the types of subjects involved in AI creation should be defined.

Firstly, the AI only undertakes the process of creating works in fact, but does not have the subject qualification due to the lack of capacity for civil rights. As defined above for the legal nature of the AI generation process, AI is based on the statutory license use of pre-existing data, the completion of machine learning and neural network model training, in order to achieve the autonomous generation of content through the adjustment of cues input. In other words, the completion of the creation of the work is the process of human input instructions, AI model execution algorithms, and ultimately get the output, i.e., a complete human-AI collaboration process. The human's ideas such as selection and arrangement in the creation process are embodied in the short cues, and finally the generated content is output from the AI algorithm model. There is no doubt that humans and AI have in fact

accomplished creation together, and thus AI has certainly undertaken the factual creation of content [3].

However, based on the completion of the factual creation process, it is inferred that AI and natural persons share the qualification of creative subject, which undoubtedly impacts the basic theory of civil subject. The essence of AI lies in its high degree of imitation of human intelligence. Its input of data, operation of algorithms, and output of content essentially imitate human learning, creative conception, and work completion [4]. The fundamental difference is that AI requires humans to continuously provide data input, technology support, and cue word interventions. This results in a passive state where AI is unable to carry out its creation without human instructions. Therefore, the autonomous generation of AI remains conditional and passive. The practice of a subject transforming the world involves infusing the subjectivity of the self into the object. The transformed object is fused with the subject's will, personality, and dignity-creation assumes the objectification of the subject's will. Human creation accomplishes works, and the process of creating works accomplishes human subjectivity [5]. In this sense, the involvement of AI cannot undermine the traditional subject-object dichotomy of creation theory. It cannot undermine the subjective status of human creation nor share the qualification of civil rights subject with natural persons. AI merely serves as a medium between the creative subject and the work. Compared to the status of natural persons as creative subjects, AI should be considered as a tool for human creation rather than a creative subject [6].

Secondly, the process of autonomous generation of AI involves three subjects: the developer, user and investor of AI.

It is currently impossible to prove the status of creative subject of AI. Therefore, it is important to define the types of subjects involved in the creation of AI. The completion of AI creation not only relies on the one-time human-AI collaboration between AI and its users, but also attributes to the designer's model training based on the pre-existing data, as well as the investor of AI product development and promotion. Among them, the AI user is the direct subject of content generation using AI tools. And in the human-AI collaborative creation model, the AI user should be categorized as the directly creative subject of the human intelligence side. The algorithm designers and investors train the model based on prior data, which plays a decisive role in the AI's ability to have mature autonomous generation. Under the human-AI collaborative creation mode, AI designers and investors should be categorized as the preparatory creation subjects on the AI side. The direct creation subject on the human intelligence side and the preparatory creation subject on the AI side jointly constructs three types of subjects for human participation in the autonomous creation of AI under the human-AI collaboration mode. Defining the subjects involved in the process of autonomous creation of AI is of great significance to the argumentation of the copyrightability of AIGC and the determination of the subject of copyright attribution.

4. Copyrightability of AIGC

The legal nature of the data-learning principle related to the process of autonomous generation of AI and the doubts related to the subjectivity has been eliminated, which provides a relatively solid foundation for the argument of the copyrightability of AIGC. Article 3 of PRC's Copyright Law defines works are intellectual achievements in the fields of literature, art and science that are original and can be expressed in a certain form. While there is little controversy in academic circles about the requirements of "literary, artistic and scientific fields" and "can be expressed in a certain form", further study is still needed to clarify the requirement of "intellectual achievement" and "originality".

AI certainly meets the requirement of "intellectual output". As mentioned in the previous two chapters, the process of AI generation is divided into five steps: prior data learning, model training, instruction input, algorithm execution and content output. Among them, data learning and model training are the preparatory creation stage, in which human intellectual creation resorts to model training and is integrated into AI in the form of algorithms. And from instruction input to the content output is the directly creating stage, in which human intellectual creation is reflected in the

arrangement, design and adjustment of cue instructions, which directly reflects participation of human intelligence [7]. AI and participation of human intelligence together constructs the AI generation and creation mode based on human-AI collaboration. The source of intelligence creation of AI is human intelligence in the final analysis [8]. Therefore, it is not difficult to prove the "intellectual achievement" requirement of AI.

The main controversy over the copyrightability of AIGC focuses on the requirement of "originality". The "originality" requirement can then be broken down into double conditions that the work is independently created and minimally creative. In other words, in order for AI to meet the requirement of originality, the process of conceptualizing, designing and completing the work should bring out the author's own unique personality. And the process should have enough space for intellectual creativity, rather than a quantitative result achieved by mechanical labor.

From the perspective of "independently created", AI users are able to derive a huge amount of feedback from sophisticated algorithms with just a few short cues, which makes their "independently created" rather questionable. The requirement of "independently created" should be analyzed in relation to the subject and the creation stage involved in the content generation. In the preparatory creation stage, the prominent contribution of AI designers and investors lies in data learning and algorithm training. At this time, the AI does not have the ability to generate content on its own. Instead, the designers and investors contribute to the development and promotion of AI products, rather than to the creation of intellectual achievement and its originality. Therefore, in the preparatory creation stage, the focus is on "preparation" rather than "creation". In the directly creating stage, the AI, which is the result of the designer's training, and its users are involved in collaborative creation. Both the human side and the mechanical side undertake the content creation process. But the autonomous generation of AI content is conditional, relying solely on the instructions provided by the user, which directly determines whether autonomous creation can take place. The user's fine-tuning of the input instructions directly determines the outcome of the output [9]. To summarize, the creativity of AIGC focuses on the selection and arrangement of cues by the AI user. In short, the creativity of the AIGC is centered on the algorithm user's original contribution to the generated content.

As far as "minimally creative" is concerned, on the one hand, it is required that it must not be exactly the same as the previous work, or have obvious recognizability with the previous work. On the other hand, it must not be quantifiable labor such as mechanical repetition. If the above two points are satisfied, it can be considered to meet the minimum creativity, which is also applicable to AIGC. AI learns through a large amount of prior data to obtain a maturely fitted algorithm. Whether it is the instruction inputted by the user to it, or the prior data and prior works used for training, the content fed back by the intelligent model after translation, arithmetic, embellishment and embellishment, is greatly different from the prior content. Moreover, even if the user repeatedly inputs the same commands to the model, these commands are translated by the algorithmic model into different machine languages for computation, which ultimately results in differentiated feedback contents. For finely tuned, selected and arranged instructions and cues, the content generated is certainly not unique and is not mechanically repetitive labor [10]. Thus, as far as the results are concerned, it is clear that the content autonomously generated by the AI is minimally creative.

In summary, the "intellectual achievement" and "originality" ("independently created" and "minimally creative") requirements of AIGC have been proved one after another. It should be considered that AIGC formally constitutes works within the meaning of copyright law.

5. Attribution of Rights in AIGC

The copyrightability of AI-generated objects has been established, making them subjects within the legal framework of copyright. Consequently, this will inevitably involve addressing issues related to the attribution of rights, specifically the subject of copyrights and the content of the copyrights. In a previous article, the subject qualification of GAI was challenged, and three types of subjects related to the autonomous generation of AI were identified. While AIGC has already taken the form of a

work (object), and the definition of the three types of subjects with original contribution is clear (subject), it is important to delve into a specific discussion regarding the scope of its legal protection (content).

5.1. Contractual Agreements Take Precedence

If there is a contractual agreement between the AI product and service provider and the user, the agreement shall be followed, as long as it does not violate the mandatory provisions of the law. Article 5 of PRC's 'Civil Code' establishes the fundamental principle of private law, meaning autonomy. Civil subjects are able to establish, change, or terminate civil legal relations in accordance with their own intentions. Chapter III of PRC's Copyright Law also outlines the content of copyright licensing use and copyright transfer contracts. As AIGC constitutes a work under copyright law, the rules of copyright licensing use and the transfer contract are also applicable. AI providers and users can stipulate the rights and obligations of both parties in the terms of service of the AI product. For AIGC that constitutes work, it can be agreed that the copyright belongs to the AI user or developer, or automatically enters the public domain upon completion of the creation [11].

5.2. Copyright Protection is Extended to the User

In other words, when the user of the AI initiates the algorithm to generate the work, their copyright is protected by the law. Looking at the subject involved in the AI creation process, the user who inputs the instructions, initiates the algorithm, and obtains the generation result is considered the direct creator of the AIGC. From the perspective of the AIGC as the object of copyright, the user's contribution to the originality of the work lies in the selection and arrangement of cues and the non-uniqueness of the generated content [8]. Furthermore, as the direct creators and disseminators of the AIGC, AI users have a dual incentive purpose of encouraging creation and dissemination, as stipulated in the first article of the Copyright Law of PRC.

It should be noted that the object of copyright recognized by law is limited to the AIGC itself. Direct creation using AI products involves three steps: instruction input, algorithm operation, and content output. Among these, the input and adjustment of instructions are only the embodiment of the user's original contribution as the creator and are necessary means and methods to complete the creation of the work. The completion of the creation and the work itself represent the essence of originality. Therefore, the object to be protected under copyright can only be the AIGC itself, not the selection and arrangement of cues. As a result, the user can only claim copyright on the AIGC itself. In addition, although the original contribution embodied in the choices and arrangements in AIGC's cues cannot be the object of the user's copyright claim, the strength of the law's affirmative protection for the AIGC itself should be commensurate with the user's original contribution. It should not expand the protection of non-affirmative claims beyond the level of the original contribution.

5.3. The Protection of Copyright Law Does not Extend to Developers and Investors

But such creative preparations are fundamentally different from the direct creation of human beings actually operating AI. Therefore, there is no basis for establishing protection for developers and investors in the copyright legal relationship. In the broader scope of civil legal relationships, such as the commercial practice of AI operators charging technical service fees to users as revenue, and AI investors receiving dividends based on their contributions and shares after product revenue, it can be seen that the legitimate rights and interests of developers and investors are still properly protected [12]. Hence, there is no need to establish protection in the form of copyright law.

6. Conclusion

The development of generative GAI brings a series of new challenges to the legal system of copyright. The correct way to address these challenges is first clarify the legal nature of the process of autonomously generating content by AIGC, focusing on the significance of AIGC in terms of the

subject, object, and content in the copyright legal relationship. A mature AI algorithm can only be developed through the developer's legally licensed use of pre-existing data, followed by model training. Subsequently, through AI user input instructions, the algorithm is activated to generate output results, leading to the final content generated by the AI. In this process, the use of pre-existing data under statutory licensing and model training constitutes the preparatory creation stage, while the input instructions, algorithms activation, and output content form the direct creation stage. Based on the theory of civil subject, the technical nature of AI working principles, and the human-AI collaboration, AI cannot be considered the subject of copyright, even if it is involved in the factual creation process. Instead, its developers, investors, and users are the key subjects related to AIGC. Human intelligence always plays a role in the process of autonomous content generation by AI, and the users, in particular, make significant original contributions to AIGC, demonstrating minimal creativity when compared with prior works. Therefore, AIGC formally constitutes a work within the scope of copyright law and is an object in the legal relationship of copyright.

Considering the human-AI collaborative creation mode and the interaction between AI and human intelligence at both ends, the user of AI emerges as the subject closet to the direct creation stage, asserting copyright over AIGC for legal protection. While the rights of AI developers and investors should not fall under copyright law, contractual agreement among these parties should take precedence based on the principle of autonomy in private law. It is important to note that the protection of AIGC users' rights under copyright law should only be limited to the AIGC itself and should be proportionate to the user's contribution to the originality of the work generated. The scope of copyright protection should be extended beyond the object or the user's original contribution to the work. In facing the challenges posed by GAI to the copyright legal system, it is crucial to analyze the legal nature of AIGC based on the subject, object, and content in the copyright legal relationship. This analysis aims to align the copyright framework with emerging technologies like GAI and ensure compatibility between intellectual property law and technological advancements.

References

- [1] Fang Huiying. Legal protection of artificial intelligence-generated content. *Shandong Social Sciences*, 2023, (09): 179-184.
- [2] Liu Shaojun, Nie Linfeng. Debate on copyright law of artificial intelligence generated content. *Journal of Nanchang University (Humanities and Social Sciences)*, 2024, 55(01): 107-118.
- [3] Zhang Jinping. On the copyright ability and copyright infringement liability of the creation of AI. *Social Sciences in Nanjing*, 2023, (10): 77-89.
- [4] Yang Lihua. Research on the copyright of artificial intelligence products. *Modern Law Science*, 2021, 43(04): 102-114.
- [5] Georg Wilhelm Friedrich Hegel. *The phenomenology of spirit*. Cambridge: Cambridge University Press, 2017.
- [6] Liu Wei. Challenge of and counter measures for copyright law in the age of artificial intelligence. *Journal of Shanghai Jiaotong University (Philosophy and Social Sciences)*, 2021, 29(02): 40-49.
- [7] Wang Guozhu. Humanistic logic in the copyrightability determination of artificial intelligence products. *Journal of East China Normal University (Humanities and Social Sciences)*, 2023, 55(01): 133-142+205.
- [8] Xu Jiali. On the copyright ownership of artificial intelligence products. *Jinan Journal (Philosophy & Social Science Edition)*, 2023, 45(04): 37-49.
- [9] Li Yang, Tu Teng. On the criterion for the copyrightability of generative AI works. *Intellectual Property*, 2024, (01): 68-84.
- [10] Jiang Ge. On the copyability of artificial intelligence generated content: from the perspective of user's original expression. *Intellectual Property*, 2024, (01): 36-67.
- [11] Tang Yili, Niu Sihan. On the subject of rights of works generated by artificial intelligence and its copyright attribution. *Fujian Tribune (The Humanities & Social Sciences Monthly)*, 2023, (11): 107-122.

- [12] Qi Aimin, Yang Yu. The evolutionary of the right confirmation mode of artificial intelligence generated content and the establishment of data property right. *Journal of Soochow University Philosophy & Social Science Edition*, 2024, 45(01): 88-99.