Proceedings of the 3rd International Conference on International Law and Legal Policy
DOI: 10.54254/2753-7048/69/20240143

The Future of Generative AI: Copyright and Infringement Issues

Jiaji Xing[1,a,*]

[1]*Beijing No.15 High School, Yuxin Street, Beijing, China*
a. xingjiaji-personal@outlook.com
**corresponding author*

Abstract: Rapid advancements in generative artificial intelligence (Generative AI) have sparked a global revolution, ushering in a new era of accelerated automation that promises to upend industries, boost productivity, and potentially displace significant portions of the white- and blue-collar workforce. However, the rise of Generative AI also presents numerous challenges, particularly in the realm of copyright and infringement. The use of copyrighted works as training data for AI models has sparked fierce debates among creators, artists, writers, and programmers who object to their creations being utilized without attribution or compensation. Whether AI-generated content should be protected under copyright law is contentious, as the traditional copyright framework was designed for human creators. This paper delves into the copyright and infringement challenges posed by Generative AI, examining the legal framework, case studies, and potential solutions. It argues that a strategic approach is necessary to navigate the complexities of this rapidly evolving technology while ensuring the protection of intellectual property rights. The paper concludes by advocating for the development of comprehensive legal frameworks that balance the benefits of Generative AI with the need to safeguard creators' rights.

Keywords: AI-generated content, Copyright, Infringement, Training data.

1. Introduction

On February 15, 2024, OpenAI, an American AI research company, officially released the artificial intelligence video generation model Sora,[1] marking the entry of AI technology into a new stage. This means that anyone who inputs text, called a "prompt," into the model can obtain video content that closely matches the description, making it difficult to identify flaws. Many film industry professionals have expressed concern and unease, stating that this will greatly reduce the difficulty of filmmaking and shorten the production cycle, which could have a disruptive impact on the traditional film industry. However, some believe that technological change is an inevitable trend in history and that embracing change and actively innovate is the the appropriate mindset and methodology.

The development of generative artificial intelligence technology has entered a fast lane, signalling the arrival of a technological revolution. Along with this revolution, several issues have arisen. Can movies, music, and literary works created by artificial intelligence be freely used? Are they protected by copyright law? Many experts and scholars have debated this issue, with some arguing that works created by artificial intelligence should be protected by law, while others hold the opposite view.

DOI: 10.54254/2753-7048/69/20240143

Meanwhile, governments in many countries are formulating relevant policies aimed at either restricting or promoting the development of generative artificial intelligence and clearly defining the protection and ownership of these works.

2. AI Copyright Issues

2.1. Determining AI Infringement and Legal Responsibility and Liability

2.1.1. Determination of AI-related Infringement: Considering the Machine Learning Process as a Case Study

In 2023, The New York Times sued OpenAI and its investment company, Microsoft, for allegedly copying nearly a million articles from its newspaper for machine learning purposes without permission.[2] According to The Hollywood Reporter, this is not the first case of its kind. Eight newspapers in the United States have already filed lawsuits against OpenAI and Microsoft with similar allegations.[3]

The discussion on the copyright of artificial intelligence is still ongoing. On March 15, 2024, the Japanese Cultural Agency released its General Understanding on AI and Copyright in Japan.[4] The document suggests that simply waiting for case law to accumulate in the absence of any precedents, the growing opposition to these issues will only exacerbate the situation, which is not conducive to the development of the industry. Therefore, it is necessary to address the concerns of all parties involved as soon as possible and provide a clear response to the copyright issues related to generative AI, enabling people to exercise better control over risks when engaging in relevant activities.

In these cases and documents, common issues can be identified. The process of AI machine learning inevitably requires the use of a large amount of material, including works and content protected by copyright. If machine learning cannot be performed, the iteration of AI becomes unfeasible. In the context of machine learning, the replication and modification of works raise legal questions. According to US Copyright Law, Title 17 U.S.C§107,108,109, the reproduction of no more than one copy or phonorecord of a work by library or archive staff for nonprofit purposes does not constitute copyright infringement. Conversely, Japanese Copyright Law stipulates that reproduction through an “automated reproduction device” or attempts to circumvent technological protection measures constitute infringement. It is evident that copying, adapting, and similar actions in machine learning are considered illegal, necessitating that relevant parties assume legal responsibility. In addition to these issues, the process of generating artificial intelligence poses additional potential risks of copyright infringement. During the creation of AI-generated works, models and algorithms process data, extracting information for experiential storage. This raises concerns about potential infringement of copyright owners’ rights to protect work integrity. Specifically, if AI generates a work using data that is similar or identical to the materials used in machine learning, does this constitute copyright infringement? These are critical issues that require careful consideration.

In light of the aforementioned cases, current legislation does not consider AI machine learning to fall within the purview of fair use. Unless legal amendments are made, such conduct remains unlawful and entails legal responsibility.

2.1.2. Legal Liability Determination and Bearing: Considering AI-Generated Content as a Case Study

Given the current scarcity of adjudicated cases of AI infringement and the limited experience or precedents available to judges and courts, establishing liability for AI infringement is imperative. The precision and equity of judgments have emerged as pivotal concerns.

DOI: 10.54254/2753-7048/69/20240143

To address this issue, it is reasonable to refer to cases from the Guangzhou courts [5] for analysis. In one case, the court ruled that, in the event of proven infringement, the defendant must cease the infringing conduct and ensure that its services no longer generate content related to Ultraman. "The degree of prevention should be such that a normal user who uses keywords related to Ultraman cannot generate images substantially similar to the works at issue," the judgment stated.

In this case, the defendant AI company should bear all legal liabilities because it failed to implement keyword filtering and other preventive measures to avoid copyright infringement. Although individuals and groups using AI services are not legally responsible, their lack of legal awareness and understanding of potential infringement indirectly contributed to the occurrence of this case. However, this ruling also raises new issues. If registered intellectual property is legally protected, does prohibiting the use of AI services to alter and create based on it hinder innovation? Legislative, executive, and judicial authorities in various countries should seriously consider this issue and reasonably define infringement.

It is evident from this case that the ruling considered a multitude of factors. In the context of judicial practice, determining copyright infringement typically requires satisfying two prerequisites: 'access' and 'substantial similarity.' The images generated by the Tab website closely resemble the original and creative expression of the artistic image of Ultraman works, exhibiting a high degree of similarity in several key features. This demonstrates substantial similarity. This finding aligns with the previously mentioned criteria for assessing the similarity between generated and original works, suggesting that these criteria could serve as a valuable framework for quantification and refinement in judicial trials.

2.2. The Copyrightability and Ownership of AI-Generated Works

2.2.1. The Copyrightability of AI-Generated Works

Recently, American artist Kristina Kashtanova created a science fiction comic book titled *Zarya of the Dawn[6]*, part of which was produced using an artificial intelligence platform called Midjourney based on Kashtanova's instructions. The author initially registered the book with the U.S. Copyright Office. However, after learning from Kashtanova's social media post that AI had been used to create part of the book, the U.S. Copyright Office refused to register the work, stating that the law only protects "works of authorship by humans." If a work is not created by a human, the Copyright Office will refuse to register it.

After further review, the U.S. Copyright Office reversed its decision and issued a new ruling on February 21, 2023. The new ruling allowed for the full registration of Zarya of the Dawn but narrowed the scope to explicitly exclude the portion generated by artificial intelligence technology (i.e., the images generated by Midjourney based on Kashtanova's prompts). The new scope of registration covers only the "author's creative words and the selection, coordination, and arrangement of AI-generated works" that Kashtanova formulated while writing the book. Images generated automatically by Midjourney are not protected.

One of the factors considered by the U.S. Copyright Office in making the above decision was the degree of human involvement in the creative process. For the images generated by Midjourney, Kashtanova only provided targeted prompts and parameters for the AI system. This did not demonstrate that Kashtanova had sufficient control over the output of Midjourney's results to qualify as the author (or co-author) of these AI-generated works. The Office determined that these images were not creative, artificially created works of art and, therefore, were not eligible for copyright protection. Based on this, the U.S. Copyright Office rejected the idea that copyright protection could be granted to images generated using artificial text prompts, stating that users of these AI systems do

DOI: 10.54254/2753-7048/69/20240143

not have sufficient human involvement or control over the creation of the images and, therefore, cannot be protected by copyright law.

The standard for human authorship is also exemplified in another case. Computer scientist Dr. Stephen Thaler used an AI program known as the "Creativity Machine" to produce a two-dimensional image titled "A Recent Entrance to Paradise" and subsequently sought copyright registration from the U.S. Copyright Office[7]. However, due to the absence of human authorship in this work, the application was not approved.

Hence, it can be inferred that the copyright eligibility of AI-generated creations depends on the extent of human involvement, and the manifestation of originality in such works must be under human control. Meeting this criterion can substantiate claims to copyright protection. In this context, it is evident that human creativity plays a pivotal role in the recognition of copyright. Similarly, in China, the Copyright Bureau and other institutions have implemented comparable criteria, suggesting a certain level of consensus between China and the United States on this issue. Granting unrestricted copyright protection to AI-generated works would significantly diminish the value of original creation, potentially leading to adverse outcomes such as the proliferation of low-quality content. With these cases providing support, the criteria for future judgments will become increasingly defined, necessitating concerted efforts from the government.

2.2.2. AI Works Copyright Ownership

In May 2017, a poetry collection written by Microsoft Bing, titled The Sun Lost Its Glass Window, was officially published[8]. This collection was created after Bing studied modern poetry from 519 poets and underwent over 10,000 training sessions. Works generated by artificial intelligence closely resemble those produced by humans, making it challenging to discern their non-human origin. Compared to previous technological advancements, the impact of artificial intelligence on copyright law is both profound and far-reaching.

Three primary proposals address the rights to works generated by artificial intelligence. The first proposal is to recognize these works as legitimate but not provide protection for them, effectively placing them in the public domain. This approach is based on the idea that copyright law aims to incentivize human creativity and dissemination, which is unnecessary for machines. The second proposal suggests establishing a new neighbouring rights system to distinguish between AI-generated works and those created by humans. The third proposal involves making appropriate legal arrangements through interpretation within the existing copyright law framework. Opinions are divided on whether copyright should belong to the owner, developer, or user of the artificial intelligence.

In the case of "Tencent's lawsuit against a Shanghai-based technology company"[9], adjudicated by the Nanshan District People's Court in Shenzhen, Judge Luo Jing noted that according to Article 9 of the Implementation Regulations of the Copyright Law of China, the current legal framework explicitly specifies that copyright can be enjoyed by citizens, legal persons, or other organizations. The law recognizes individuals (including natural persons and legal entities such as corporations and other organizations) as the only entities with rights and obligations. Consequently, AI itself cannot claim copyright protection since it is not a recognized legal entity. Therefore, the copyright for original works generated by AI should belong to the developer of the program.

However, in exceptional circumstances, the determination of copyright ownership for AI-generated works must be based on the specific context. According to Luo Jing, if an individual is the developer, they can be considered the copyright owner; if a legal entity or other organization is responsible for the development, then they should be regarded as the copyright owner. The specific copyright ownership should also consider the level of contribution by the developer and any contractual agreements between them and investors. If the user agreement for an AI application

DOI: 10.54254/2753-7048/69/20240143

explicitly outlines copyright ownership, then users who generate a paper using this application and subsequently sign and publish it would indeed infringe upon the copyrights held by the AI software developer.

3. The Necessity and Value of Copyright Protection for Artificial Intelligence Works

In the contemporary era, AI technology is advancing rapidly, and its pivotal role in both production and daily life is poised to become increasingly significant. Furthermore, the quality of AI-generated works is expected to undergo substantial improvement. It is widely acknowledged that machine learning plays a crucial role in the development of AI. Without machine learning, the generation of AI would be unattainable. Drawing from the discourse in the preceding section, it is imperative to include machine learning-generated content within the purview of fair use, and it is encouraging to observe that several countries and regions have already implemented this proposal. The European Union, the United Kingdom, France, Germany, Japan, and South Africa have incorporated machine learning mining materials, whether limited or unlimited, into the realm of fair use. This will significantly advance the progress of AI technology, yielding substantial productivity gains for these countries and regions.

Following the resolution of the machine learning problem, the subsequent critical topic for consideration pertains to the copyrightability of AI-generated content. Although it has been previously noted in this essay that determining copyright licensing for AI-generated works is stringent, there are instances where AI-generated works have obtained copyright protection. This underscores the significant implications for safeguarding them from diverse perspectives. In essence, safeguarding AI-generated creations equates to safeguarding the advancement of innovation. AI-generated comic books, short films, and novels are increasingly gaining prominence, these creative works are poised to significantly enhance our cultural landscape and cater to the demands of our intellectual and emotional growth.

From an ethical standpoint, granting copyright protection to human-AI collaborative works is justifiable. Traditional ethical principles have historically confined legal subjectivity to humans, excluding animals and other living entities. Granting copyright protection to AI does not violate these ethical foundations; rather, it upholds human worth by acknowledging the contributions of humans in the creation process, rather than equating robots with humans in legal terms. Therefore, extending copyright protection to AI-generated works does not conflict with established ethical norms regarding humanity.

In terms of social security, it is essential to grant copyright protection to AI-generated works. As the AI industry advances, ensuring that AI development remains within manageable bounds is crucial for human safety. Granting copyright to AI-generated works helps promote the healthy growth of the AI industry within legal parameters and prevents misuse and disorder. However, when conferring copyright on AI-generated works, it is important to elevate the standard of “originality” expected from these works. This standard may need to exceed traditional criteria for originality significantly. Works created solely through reliance on models may be excluded from copyright protection if they fail to reflect genuine creative input from an author.

4. Proposals

From the aforementioned legal amendments, certain challenges remain in integrating machine learning into fair use. Specifically, these amendments exhibit a tendency towards conservatism. The revised laws predominantly cover non-profit, research, and educational purposes, as well as other highly restricted scenarios within the realm of fair use. Consequently, it can be inferred that the

DOI: 10.54254/2753-7048/69/20240143

widespread application of free fair use does not align with the best interests of copyright holders. This situation clearly hinders the advancement of artificial intelligence technology.

Given the anticipated reliance of future artificial intelligence on the extensive use of machine learning materials, a centralized and unified management approach may prove to be an efficient strategy[10]. The powers of this organization should be clearly delineated, including the establishment of licensing agreements with users, the collection and distribution of royalties to rightsholders, and the facilitation of litigation and arbitration proceedings. Additionally, to streamline authorization and licensing procedures and enhance the operational efficiency of collective management organizations, it may be beneficial to transcend the constraints related to subject matter and work types by establishing specialized copyright collective management organizations for works used in algorithm training. It is essential to establish uniform charging standards and distribution criteria based on work usage. Furthermore, continuous improvement in the technical expertise and managerial capabilities of collective management organizations is crucial for advancing the standardization and efficiency of large-scale market utilization authorization practices.

Moreover, the simultaneous implementation of regulatory measures for AI use is essential. When considering control measures for AI usage, the primary focus should be on assessing whether subsequent avoidance behaviour constitutes infringement and whether avoidance based on reasonable use qualifies for an exemption[11]. In the future, the avoidance of implementing control measures can be justified from a systemic interpretation perspective by legitimate use. Additionally, apart from instances where technical measures and legitimate use coincide under statutory circumstances, other scenarios of legitimate use, encompassing both statutory and general situations, should be comprehensively considered. This entails taking into account contextual factors such as means, methods, and significance, and undergoing the “three-step test.”

5. Conclusion

Looking ahead, the advent of artificial intelligence is imminent. The integration of AI will lead to a promising future and that the continual advancement of technology will undoubtedly yield substantial benefits. Meanwhile, devising strategies to address the ensuing opportunities and challenges is a pivotal consideration for all.

However, as artificial intelligence progresses, a spectrum of challenges is anticipated to arise. The patentability of AI-generated works has significant implications and value, fostering advancements in this new era while also presenting novel challenges. If AI-generated works are widely granted copyright, does the concept of copyright lose its significance? While the function of copyright is continually evolving, it remains of considerable significance. In the foreseeable future, the presence of copyright will better serve the interests of society.

References

[1] *OpenAI. (2024, February 15). Sora | OpenAI. https://openai.com/index/sora/*

[2] *Duffy, C., & Goldman, D. (2023, December 27). The New York Times sues OpenAI and Microsoft for copyright infringement | CNN business. CNN. https://edition.cnn.com/2023/12/27/tech/new-york-times-sues-openai-microsoft/index.html*

[3] *Cho, W. (2024, May 1). Major daily newspapers join legal fight against OpenAI as battle lines are being drawn. The Hollywood Reporter. https://www.hollywoodreporter.com/business/business-news/newspapers-lawsuit-openai-1235887081/*

[4] *Japanese Cultural Agency. (2024, March 15). General Understanding on AI and Copyright in Japan. Cultural Agency. https://www.bunka.go.jp/seisaku/bunkashingikai/chosakuken/bunkakai/69/pdf/94022801_01.pdf*

[5] *XinBang. (2024, February 26). Exclusive: AI Creates Ultraman: Chinese Court Makes First Effective Judgment of Generative AI Service Infringing Copyright Globally. 21Economic.com - Official Website of 21st Century Economic Report. https://www.21jingji.com/article/20240226/herald/133a6c2f9c0b045899e4dea10c5778eb.html*

DOI: 10.54254/2753-7048/69/20240143

[6] *Kris, K. (2023, March 9). (AI) Zarya of the dawn - Midjourney — Kris Kashtanova. Kris Kashtanova. https://www.kris.art/portfolio-2/project-one-ephnc-jamy8*

[7] *Mattei, S. E. (2024, January 5). Stephen thaler's quest to get his 'Autonomous' AI legally recognized could upend copyright law forever. ARTnews.com. https://www.artnews.com/art-in-america/features/stephen-thaler-quest-ai-legally-recognized-upend-copyright-law-1234692243/*

[8] *Chat bot's poetry irks scholars. (2017, June 8). The Poetry Foundation. https://www.poetryfoundation.org/poetry-news/77504/chat-bots-poetry-irks-scholars*

[9] *Ning, Z. (2023, June 8). Tencent Co., Ltd. v. Shanghai Yinxun Technology Co., Ltd 2018 Copyright Infringement Lawsuit. Flatfee. https://flatfeecorp.com/pt/articles/tencent-versus-shanghai-yinxun-technology-2018-copyright-infringement-case-china*

[10] *Qingwen, L. (2024). Algorithm training utilizes the legal framework of copyright law to regulate the use of works. Science and Publishing, (7). https://doi.org/10.16510/j.cnki.kjycb.20240719.003*

[11] *Wenkang, Z., & Yanying, F. (2024). Generative artificial intelligence creations employ fair use adaptation of existing works. Science Technology and Law (Chinese-English Version), 3. https://doi.org/10.19685/j.cnki.cn11-2922/n.2024.03.009*