# AIGC (Artificial Intelligence Generated Content) infringes the copyright of human artists

#### Lyulin Zhuang

Department of Maynooth International Engineering, Fuzhou University, Fujian, 350000, China

lyulinzhuang@163.com

Abstract. With the rapid development of artificial intelligence technology, content generated by artificial intelligence has been rapidly applied to people's lives. At the same time, it is also accompanied by many infringement lawsuits, whether AIGC has really caused different degrees of infringement to human artists. Through the analysis of the existing literature on copyright issues and the walkthrough of Stable Diffusion, an AI-generated image platform, this article digs into the main factors that the AI-generated platform causes infringements on human artists. Provide references for using AI by enterprises and related media, and let more scholars pay attention to this issue. The study found that in the workflow of the AI generation platform, taking Stable Diffusion as an example, the two processes of model training and image generation may cause copyright infringement to a certain extent. Based on this, the AI generation platform has unauthorized use of copyright works, excessive plagiarism and adaptation of copyright works, and the generated images are not marked with watermarks or sources, which damages the copyright owner's rights.

Keywords: AIGC, Copyright, Human Artists, Stable Diffusion.

#### 1. Introduction

The emergence of artificial intelligence-generated content (AIGC) has brought significant advancements in various fields, such as advertising copywriting, news reporting, and scientific creation. This technology, known as AIGC, has led to reduced labor costs and improved creative efficiency, thereby delivering greater commercial and creative value to businesses and creators. However, controversies surrounding copyright issues and innovative capabilities during the content generation process have become a subject of concern.

As a generative AI technology for text-to-image generation, the Stable Diffusion technique has garnered global attention. While this technology appears remarkably innovative, it also brings forth certain issues, including copyright concerns.

Stable Diffusion can quickly generate high-quality images based on simple text descriptions, thanks to its training on billions of sample images collected online. However, not all of these sample images fall within the public domain; some are even protected by copyright. As one can imagine, the authors and photographers of these copyrighted images are not pleased with the actions of Stable Diffusion.

Stable Diffusion has faced multiple lawsuits due to copyright issues. Artists and photographers have initiated collective lawsuits against this technology, and one of the world's leading image suppliers,

Getty Images Holdings Inc., has also filed a similar case. According to Getty's (2023) complaint, "Stability AI has reproduced over 12 million Getty Images, along with their related captions and metadata, without permission or compensation." Consequently, delving deeper into this issue and exploring potential solutions is essential.

AIGC model training phase. One critical aspect of using copyrighted works in AIGC is the two-stage process of utilizing these works. To illustrate this process, the Stable Diffusion technique will be used as an example. During the model training phase, copyrighted works are copied from LAION-5B (Large-Scale Artificial Intelligence Online) database and modified before being incorporated into the image information space. This process ensures convenient output for future user consumption.

AIGC model output stage. However, the second stage of output generation raises concerns regarding the potential infringement of copyrighted works. If the generated image is deemed "substantially similar" to the original work in expression, it may infringe upon the original work's "reproduction right." Similarly, if the image forms a new word based on the original work while retaining its fundamental expression, it may infringe upon the original work's "adaptation right."

With the widespread adoption of AIGC, numerous original authors have resorted to legal actions against commercial companies utilizing AIGC technology. This trend indicates a growing concern for copyright infringement incidents in AIGC.

While considerable progress has been made, there is still a long way to go in finding a harmonious balance between AIGC development and artists' rights. Based on this, the study aims to contribute to this ongoing endeavor by undertaking more in-depth research and comprehensively understanding the issue at hand.

As AIGC continues to evolve and revolutionize creative industries, it is imperative to address copyright concerns and find suitable solutions that safeguard the intellectual property rights of artists while embracing the benefits offered by AIGC technology. This study is designed to bring more attention to the other side of AIGC than just the convenience it gets so that people can use this technology with respect for the fruits of other people's work and to protect the interests of more artists. By examining the literature on copyright issues in AIGC technology, we aim to delve into the intricacies of this field's challenges and potential solutions, providing valuable insights and guidance for academic research and practical applications in related domains.

# 2. Methods

#### 2.1. Participant

The participation of this study revolves around the images generated by AIGC. The images generated by AIGC refer to computer-generated images, graphics, or visual content produced through artificial intelligence. These images are typically generated using AI algorithms and deep learning models. They can be completely new images generated from scratch or modified and edited based on existing ones. Taking Stable Diffusion as an example, users can generate images using pure prompts or prompts combined with images.

The images generated by AIGC can have various forms and styles, ranging from realistic photo compositions to abstract artworks. As Marek J. Magdalena Z. (2023) mentioned, AI-generated images are created by training models on vast amounts of data and using that data to generate novel visual content [1]. As Brownlee J. (2019) published, the algorithms involved in image generation may utilize convolutional neural networks (CNNs), generative adversarial networks (GANs), or other machine learning methods [2].

AIGC-generated images find application in numerous fields, including visual effects production in the film and gaming industry, artistic creation, product design, and virtual reality. However, due to potential copyright and originality issues associated with AIGC-generated images, their usage and protection have sparked numerous copyright disputes and legal discussions.

Due to the substantial increase in copyright infringement cases involving AIGC-generated images compared to other AIGC-generated content in recent years and the widespread use of AIGC image-

generation technology in various industries, the risk of copyright disputes is higher. Furthermore, images are more susceptible to infringement and can be easily identified by copyright holders, compromising their interests. Therefore, it is justified to choose AIGC-generated images as the research subject.

# 2.2. Literature Review

This study uses various secondary data research tools, including academic journals, research reports, industry news, and insights and comments from relevant experts and scholars. Multiple educational websites such as gizmodo.com, arxiv.org, and techcrunch.com were utilized to access professional research literature. Academic search engines such as Google Scholar, domestic Q&A platforms, and social media platforms like Zhihu and Xiaohongshu were employed to search for online resources, gain inspiration, and engage in academic discussions. The collected secondary data was stored as electronic files on laptops and iPads.

Firstly, the secondary data was read and comprehended using software such as MarginNote3 and Google Translate to grasp key concepts, theories, or perspectives and understand the authors' research methods and findings. Secondly, the read secondary data were categorized and summarized. Utilizing annotations, excerpts, or summaries, essential information from the literature was extracted and synthesized into a collection of relevant concepts, models, or theories.

Then, the synthesized and summarized secondary data was compared and analyzed. Different literature sources often offer varying perspectives and emphases on the same issue—in-depth analysis of diverse secondary data aids in gaining a comprehensive understanding of the research topic.

Literature synthesis and analysis are commonly used research methods for secondary data. These methods were chosen due to the limitations of time and resources, which barred conducting field surveys and experimental research. Secondary data provides a wealth of research outcomes and academic discussions, enabling deep exploration of copyright issues in AIGC technology and an understanding of various aspects and potential impacts within this field.

A large body of academic research findings can be integrated and synthesized through literature synthesis to construct a comprehensive research framework while organizing and managing secondary data through categorization and classification. Literature analysis helps in-depth reading and understanding of relevant literature, extracting critical information, and forming new insights and experiences through comparison and analysis. These methods provide a more comprehensive understanding of the research object and a theoretical and empirical foundation for subsequent research and discussions.

# 2.3. Walk-through

The researchers adopted a walk-through research method to conduct a more specific and in-depth study. Taking Stable Diffusion as the research object, explore the working principle of the AI generation platform and study its relationship with copyright infringement.

Stable Diffusion is an AI-generated image platform widely used in the market and has also received many artists' rights protection. Therefore, using it as the research object of the walk-through can provide accurate and universal research results.

#### 2.4. Procedure

Firstly, after clarifying the general direction of researching the AIGC copyright problem, read real cases about AIGC copyright infringement cases to understand the seriousness and impact of the AIGC copyright problem. The objectives of the research were determined: to study the AIGC copyright problem, to explore solutions, to provide valuable insights and guidance for academic research and practical applications in related fields, and the research question: AIGC copyright infringement problem.

Secondly, I searched for secondary data in Google Scholar, China Knowledge Network, and other literature websites for the study. I screened secondary data that met the research objectives and questions through keywords, literature titles, and abstracts. For example, the basic principle of AIGC and the process involving infringement, existing solutions, and possible solutions to be discovered. After

collecting the secondary data, assess the reliability and applicability of the data. Factors such as research methods and research purposes are considered to ensure the reliability and validity of the secondary data.

Finally, the secondary data that is the subject of the study is identified, and then the data is extracted and organized. Articles have different research purposes for similar problems, including further research focuses. So, some articles focus on collecting data about real infringement cases of AIGC, some contain the basic principles of AIGC, and some collect the existing solutions and possible future solutions.

- Jay Alammar: The Illustrated Stable Diffusion, https://jalammar.github.io/illustrated-stable-diffusion/

- Kaixin Zhu& Yiqun Zhang: "Your AI infringes my copyright": Talking about the copyright protection behind AIGC, https://www.yuque.com/wikidesign/ykf0s9/hz2dtanx76d81661

- James Vincent: Getty Images sues AI art generator Stable Diffusion in the US for copyright infringement, https://www.theverge.com/2023/2/6/23587393/ai-art-copyright-lawsuit-getty-images-stable-diffusion

AIGC companies, taking Stability AI as an example, as the company accused of copyright infringement, are the main stakeholders of copyright issues, and they need to face legal actions, financial losses, and reputation damage.

Copyright holders must protect their works' copyright and defend their interests.

Users of AIGC, such as those of Stable Diffusion who use its generated content, need to know whether the generated content infringes copyright and whether it will impact their use.

Stable Diffusion is an AI-generated image platform widely used in the market and has also received many artists' rights protection. Therefore, using it as the research object of the walk-through can provide accurate and universal research results.

#### 3. Results

#### 3.1. How the Stable Diffusion Works

Based on the research of Alammar, J. (2022) The Illustrated Stable Diffusion. The Illustrated Stable

Diffusion – Jay Alammar – Visualizing machine learning one concept at a time. https://jalammar.github.io/illustrated-stable-diffusion/. There are some main components to run the work of Stable Diffusion [3].

In training the model, the image encoder is responsible for translating training images to information like an array of numbers and saving them.

When the user enters a prompt, the text decoder translates the prompt into an array of numbers and transmits it to the image information creator, which will run for multiple steps to generate the image. Finally, the image decoder paints a picture from the information it got from the information creator and sends it to the user [4].

To put it simply, the workflow of the stable diffusion model is like a simple chocolate-making process—collect the ready-made chocolate (Original image), melt it in the pot (Image encoder), pour the melted chocolate slurry into Containers for storage (Latent space). Then, according to the needs (Prompt) of customers (users), customize the molds (Diffusion model) according to the needs of the model factory (Text encoder), pour the melted chocolate slurry into the molds, and then perform solidification treatment (Image decoder), and finally get new chocolates (Generated image) that meet customer requirements.



Figure 1. Stable Diffusion Model

# 3.2. From what aspects does Stable Diffusion infringe the copyright of human artists

After exploring how artificial intelligence (AI)-generated platforms work, learn what processes may involve human artist infringement in the course of their work. So let's talk about where these workflows infringe on human artists' copyrighted works — Artificial intelligence (AI)-generated content raises concerns about copyright infringement of human artist works in several ways.

The researchers summarized three aspects of human artist infringement: model training using only copyrighted works without permission. Take Stable Diffusion, which has been involved in many copyright disputes. It uses the image database containing hundreds of millions—LAION—5B (Kaixin Zhu, & Yiqun Zhang. (2023)) as the source of training data, and it does not need the consent of the copyright owner [5].

Second, there is a lot of plagiarism and adaptation of copyrighted works in AI-generated images. In a case in which an artist sued an AI-generated platform, the painter Erin Hanson provided strong evidence that the AI-generated images plagiarized his work. And there are many such lawsuits now, not a few.



Figure 2. Left: Painter Erin Hanson's work in 2021

Right: Results generated in Stable Diffusion using "style of Erin Hanson" and other prompts.

Third, the images generated by the AI generation platform, such as Stable Diffusion, if works infringe copyright works, do not have any watermarks or source marks, making it even more difficult for copyright owners to accept. Unlike artificial intelligence, most human artists will mark the original author or source when learning to learn from other people's works so as not to infringe the copyright of others. Artificial intelligence often regards works as different data, lacking awe and moral measurement for works.

All in all, during the generation process of the artificial intelligence generation platform (take Stable Diffusion as an example), it may infringe on the works of human artists from two aspects of model training and image generation, and the generated results are different in three dimensions. Infringements are caused to a certain extent, including unauthorized use, excessive plagiarism and adaptation, and lack of attribution to the original author of the copyrighted work. These are the key factors that constitute the infringement of human artists.

# 4. Discussion

# 4.1. Use of copyrighted works or IP without permission

Based on a thorough examination of various critical factors contributing to the infringement of human artists' works by artificial intelligence-generated images, the researchers conducted a more focused and in-depth analysis to address this pressing issue.

Among the primary concerns is the unauthorized use of copyrighted works. To vividly illustrate this problem, the researchers used the example of Mickey Mouse and inputted the "Mickey Mouse, cooking" prompt into the Stable Diffusion algorithm (Fig.3). As we are well aware, Disney's legal department is known for its stringent control and heightened sensitivity towards protecting the image of Mickey Mouse [6]. Despite this, Stable Diffusion generated ideas that utilized Mickey Mouse's likeness for secondary creation. Additionally, the vast image databases used to train AI models present another challenge to avoiding copyright infringement. Since the training data is sourced from the Internet, it isn't easy to ensure that only authorized and copyright material, thereby exacerbating the potential for infringement. This instance highlights a significant challenge: even renowned companies with well-equipped legal teams and substantial financial resources can struggle to safeguard their copyright rights. Consequently, for many unknown artists and individuals engaged in creative work, encountering copyright infringement often leaves them feeling helpless and lacking adequate protection for their artistic creations.



Figure 3. Mickey Mouse cooking

Unlike prominent artists and corporations with the financial means and time to defend their rights, As Mark S. (2004) claimed, most unknown artists lack the necessary resources and support to protect their creative output effectively [7]. The unequal power dynamics in copyright infringement cases tend to favor large entities, leaving lesser-known artists vulnerable and their rights compromised. This glaring disparity raises legitimate concerns regarding the equitable treatment of artists in the face of AI-generated content's potential infringement on their original works.

# 4.2. Copying or modifying elements of a copyrighted work

In generating images, the artificial intelligence generation platform led by Stable Diffusion often directly plagiarizes the images used for training or moves the elements in them. In Getty v. Stable Diffusion (James V, 2023), Getty showed ample evidence that Stable Diffusion recreated its watermark in images generated by Stable Diffusion, which Getty argued appeared on models' "bizarre or grotesque images that would be blurred or defaced." and impair the quality of the Getty Images markup" [8]. , such incidents occur frequently, violating the interests of the copyright owner and quickly confusing the public, making it difficult to distinguish the actual owner of the copyright. Additionally, Companies using AI-generated technology profit from it. As Lucas R. (2023) projected, "After OpenAI freely exploited everybody's web content, it then proceeded to use that data to build commercial products that it is now attempting to sell back to the public for exorbitant sums of money." [9]

# 4.3. No guarantee of uniqueness, exclusivity

Another critical aspect of infringement is that current AI lacks self-awareness and moral judgment. In a lawsuit (Chloe Xiang, 2022), Microsoft sued the AI company; Matthew Butterick, an attorney, and programmer, says that Microsoft's offering of Copilot as an alternative to open-source code violates copyright and removes the incentive for programmers to explore open-source communities. To Butterick, Microsoft's compartmentalization of open-source code breaks the ethos of open-source programming, in which programmers often voluntarily share code as part of their mutual learning and development [10].

As Lucas R. (2023) demonstrated, "Multiple lawsuits have argued that AI companies like OpenAI and Midjourney are basically stealing and repackaging millions of people's copyrighted works and then selling a product based on those works; those companies, in turn, have defended themselves, claiming that training an AI generator to spit out new text or imagery based on ingested data is the same thing as a human writing a novel after having been inspired by other books. Not everybody buys this claim, leading to the growing refrain "AI is theft." [11] Unlike human creators who can learn from each other's techniques and ideas, artificial intelligence systems cannot comprehend and internalize the nuances of works created by human artists. Instead, they treat these works as mere data points and rely on algorithms to generate new images, often imitating the works unquestioningly. Ethical considerations are often overlooked in this process, resulting in potential issues related to the appropriation and transformation of existing copyrighted material. As Tomas G. (2023) said, "Individual artists look at each other's work and draw inspiration to create new pieces. Iterating and building on collective ideas is how art works— perfectly legal, often even blatant rip-offs. Are AI image generators just doing the same thing, or are they breaking the law?" [12]

Furthermore, the uniqueness and exclusivity of AI-generated images may be compromised due to the abundance of similar or duplicate pictures in the training data. As AI systems source data from the vast expanse of the Internet, they can inadvertently generate content that lacks diversity and originality. For instance, Stable Diffusion can repeatedly produce similar images until the user's requirements are met, leading to a proliferation of similar-looking artwork.

Another limitation lies in AI's tendency to mimic patterns and features from the training data, often producing images that closely resemble what they have seen. While the execution of AI-generated artwork can be impressive and visually striking, it may not consistently deliver entirely fresh and innovative ideas. This can be attributed to AI's inherent lack of genuine creativity and deep understanding of the subject matter, affecting the artistic novelty and originality level in the generated content.

In conclusion, the increasing prevalence of AI-generated content raises legitimate concerns surrounding the potential infringement of human artists' works. Key issues include the unauthorized use of original works for training AI models, the challenges in filtering sensitive copyrighted material, and the limitations in achieving true originality and creativity in AI-generated content.

# 5. Conclusion

This study employs literature analysis and walk-through methods to comprehensively examine the existing literature on Artificial Intelligence Generated Content (AIGC) and its implications on copyright issues. Additionally, it takes Stable Diffusion as a specific example to explore the inner workings of an artificial intelligence generation platform, establishing a solid foundation for further research in this domain. Through a meticulous investigation, the study identifies the conditions and methods by which AIGC may infringe on the rights of human artists, critically analyzing this problem in-depth.

The insights from this research offer valuable practical experience for companies intending to integrate AI technology into their products. By understanding the potential pitfalls associated with AIGC, these companies can implement measures to prevent infringement on the rights of users and artists alike. Moreover, the findings raise awareness among future scholars and researchers regarding the ethical implications of AIGC and its impact on human artists' rights.

Regarding the event, the AI-generated avatar function by Lofter was met with significant resistance from numerous users. This case study reveals that the platform undeniably infringed upon the copyrights of authors who had shared their works on the forum. The unauthorized use of these works and the generation of derivative content that closely mimicked elements of their original works amounted to a violation of their creative rights. Moreover, the generated results lacked any attribution to the original authors, further exacerbating the issue of copyright infringement.

It is important to emphasize that resistance to all AI-generated content is not necessarily reasonable. As Stephen Hawking said, "It is all right to make mistakes; nothing is perfect because without perfection, we would not exist." AI will play an increasingly significant role in creative fields as technology advances. However, addressing the challenges and potential issues arising from using AI in content generation is imperative. Recognizing that no technology is flawless, the study advocates for a balanced approach that acknowledges both the benefits and challenges associated with AIGC.

By paying attention to the multifaceted aspects of AIGC, stakeholders can foster an environment that promotes responsible AI development and usage. As the digwatch (2023) said, Industry leaders partner to encourage accountable AI development [13]. This entails ensuring that adequate measures are in place to protect the rights of content creators, preventing unauthorized use of copyrighted material, and promoting fair attribution practices in AI-generated content.

In conclusion, this research serves as a comprehensive exploration of the intersection between AIGC and copyright issues. Through a thorough analysis of existing literature and a specific examination of Stable Diffusion, the study sheds light on the potential challenges human artists face in the context of AI-generated content. It also offers valuable insights for companies seeking to leverage AI technology and calls for a nuanced and balanced approach to navigating the complexities of AIGC. Ultimately, promoting a responsible and ethical application of AI will safeguard the rights and interests of both content creators and users in the rapidly evolving landscape of digital creativity.

# References

- Marek J, & Magdalena Zemelka-Wiacek, & Michal O, & Oliver P, & Thomas E, & Maximilian R, & Cezmi A. (2023) *The artificial intelligence (AI) revolution: How important for scientific work and its reliable sharing.* https://onlinelibrary.wiley.com/doi/10.1111/all.15778
- [2] Brownlee J. (2019). A Gentle Introduction to Generative Adversarial Networks (GANs). https://machinelearningmastery.com/what-are-generative-adversarial-networks-gans/
- [3] Alammar, J. (2022). *The Illustrated Stable Diffusion*. https://jalammar.github.io/illustrated-stable-diffusion/
- [4] Guodong Zhao. (2023). How Stable Diffusion works, explained for non-technical people. https://bootcamp.uxdesign.cc/how-stable-diffusion-works-explained-for-non-technicalpeople-be6aa674fa1d
- [5] Kaixin Zhu, & Yiqun Zhang. (2023). "Your AI infringes my copyright": Talking about the copyright protection behind AIGC. https://www.yuque.com/wikidesign/ykf0s9/hz2dtanx76 d81661

- [6] Manish J. (2023) All About Disney Copyright Infringement. https://bytescare.com/blog/disneycopyright-infringement/
- [7] Mark S. (2003). *How Current Copyright Law Discourages Creative Output: The Overlooked Impact of Marketing*. https://www.jstor.org/stable/24116714
- [8] James V. (2023). Getty Images sues AI art generator Stable Diffusion in the US for copyright infringement. https://www.theverge.com/2023/2/6/23587393/ai-art-copyright-lawsuit-gettyimages-stable-diffusion
- [9] Lucas R. (2023) A New Class Action Lawsuit Adds to OpenAI's Growing Legal Troubles. https://gizmodo.com/a-new-class-action-lawsuit-adds-to-openais-growing-lega-1850593431
- [10] Chloe Xiang. (2022) GitHub Users File a Class-Action Lawsuit Against Microsoft for Training an AI Tool With Their Code. https://www.vice.com/en/article/bvm3k5/github-users-file-aclass-action-lawsuit-against-microsoft-for-training-an-ai-tool-with-their-code
- [11] Lucas R. (2023) A New Class Action Lawsuit Adds to OpenAI's Growing Legal Troubles. https://gizmodo.com/a-new-class-action-lawsuit-adds-to-openais-growing-lega-1850593431
- [12] Tomas G. (2023) Do AI Art Tools Break Copyright Laws? Two New Lawsuits Will Find Out. https://gizmodo.com/free-ai-art-stability-midjourney-deviantart-sued-getty-1849995371
- [13] digwatch (2023) Industry leaders partner to promote responsible AI development. https://dig.watch/updates/industry-leaders-partner-to-promote-responsible-ai