SCHOOL OF
ECONOMICS AND
MANAGEMENT

Copyright Infringement in AI-Generated Artworks

Lin Yin

DEPARTMENT OF BUSINESS LAW

Master's Thesis in European and International Trade Law

15 ECTS

HARN63

Spring 2024

Contents

Abstract

At a time when artificial intelligence is developing rapidly, this thesis explores the issue of copyright infringement of AI-generated artworks. This study aims to analyze the legal complexities brought about by the increasingly widespread application of artificial intelligence in the creation of original artworks, with a special focus on the protection of creators' rights and intellectual property rights. Through an in-depth analysis of relevant cases and legislation in the United States and the European Union, combined with the concept of reproduction rights and exceptions and limitations to copyright, this thesis explores the impact of AI-generated works and technologies such as machine learning on the exclusive rights of rightholders. Many issues regarding generative AI and copyright are still unclear in judicial practice, but machine learning may indeed infringe on the right of reproduction. It is also very important to find a balance between copyright protection and ensuring space for technological development.

Keywords: copyright infringement, AI-generated artworks, intellectual property law, European law, fair use, exceptions and limitations to copyright

Abbreviations

AI	Artificial Intelligence
ANN	Artificial Neural Network
CJEU	Court of Justice of the European Union
EU	European Union
GPAI	General-purpose Artificial Intelligence
NN	Neural Network
SNN	Spiking Neural Network
TDM	Text and Data Mining
TRIPS	Trade-Related Aspects of Intellectual Property Rights
US	United States
WCT	WIPO Copyright Treaty
WTO	World Trade Organisation
WIPO	World Intellectual Property Organization

1 Introduction

1.1 Background

In recent years, AI (artificial intelligence) has developed at an amazing speed. Since the launch of ChatGPT[1], ("a prototype dialogue-based AI chatbot capable of understanding natural human language and generating impressively detailed human-like written text"[2]) generative AI including various generative language models and generative image models, which brings challenges to many areas. Some people are worrying that generative AI would replace human creators, including academia, literature, art, journalism and many other fields.[3]

In fact, in the field of illustration, generative AI has already had some impact on human creators. In 2022, AI-generated artworks began appearing on ArtStation, one of the most famous illustration platforms. From then on, artists began their series of protests against those AI works. They believe that both the act of AI training and the final output of AI are unethical and even infringe their copyright exclusive rights.[4]

During the training process, works protected by copyright are used in AI training without rightholders' permission. While AI training process involves the producing copies and storage of copyrighted works, this may infringe the exclusive rights of the copyright owners. As for the final output works, some of them retain traces of the original works. For instance, they may have similar structures, colors, and lights and shadows with the original ones, etc. In some circumstances, you can even see traces of a watermark or signature. In response to this, ArtStation removed the images protesting AI-generated art from its homepage and rejected the their request.[5]

After that, the artists' protests have not stopped. In January 13th 2023, on behalf of themselves and other artists, Sarah Anderson, Kelly McKernan and Karla Ortiz filed a putative class action against three AI companies including Stability, DeviantArt and Midjourney in California for copyright infringement. According to Courtlistener website, this case has not been settled yet.[6] This case is highly representative and its verdict will have a significant impact on the development of AI-generated artworks in the United States.

---

[1] See Openai website: https://openai.com/

[2] Samantha Lock, 'What Is AI Chatbot Phenomenon ChatGPT and Could It Replace Humans?' (The Guardian5 December 2022) <https://www.theguardian.com/technology/2022/dec/05/what-is-ai-chatbot-phenomenon-chatgpt-and-could-it-replace-humans>

[3] See ibid.

[4] Game Corner, '"Artist Collective Protest against AI Drawings Occupied ArtStation Homepage, Epic Games Failed to Appease It, but Became More and More Violent"' (United News Network, 3 January 2023) <https://tech.udn.com/tech/story/123158/6840587> accessed 23 May 2024

[5] See Jess Weatherbed, 'ArtStation Is Hiding Images That Protest against AI Art' (The Verge23 December 2022) .

[6] See 'Andersen v. Stability Ai Ltd., 3:23-Cv-00201 – Courtlistener.Com' (Courtlistener) <https://www.courtlistener.com/docket/66732129/andersen-v-stability-ai-ltd/> accessed 23 May 2024

There are also some other cases in the US (United States) about generative AI, most of these cases are still under trail. In these cases, when the plaintiff claimed that the defendant's unauthorized use of their works to train AI models infringed the copyrights, the defendant basically cited fair use as a defense.[7]

The most relevant exclusive rights of copyright to AI training is reproduction right. However, the definitions of reproduction is different in different jurisdictions. Whether AI learning can be attributed to the definition of reproduction right is still uncertain in international treaties and domestic laws of various countries. This is also an important reason why the infringement of AI training has caused controversy. If such use does not infringe any exclusive rights, then all other issues are irrelevant.

The definition of Berne Convention covers a wide range of reproduction rights, and it can be considered that it covers machine learning and AI training. However, in US, this determination may have different standards in different courts. Compared with them, the EU (European Union) is more clear in its definition.[8]

However, if the use of data for training AI meets the definition of reproduction right. Whether this use complies with the provisions of the exceptions and limitations of copyright exclusive rights becomes the key to answering this question.

Intellectual property rights are not always an absolute right including copyright. They have to be restricted by many conditions. In Article 107 of the copyright law of the United States, the fair use of a copyrighted work, "including such use by reproduction in copies or phonorecords or by any other means specified by that section, for purposes such as criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research, is not an infringement of copyright".[9]

The United States does not have specific legislation on copyright issues regarding artificial intelligence, but there are more and more related cases in practice, and there are still many uncertainties in current judicial practice. Compared with the United States, the EU is relatively backward in the practice of artificial intelligence technology, but it is ahead in legislation.

In EU law framework, Article 5 of Directive 2001/29/EC, exceptions and limitations are provided.[10] Directive (EU) 2019/790 also introduced rules of copyright exception and data mining. These rules have similarities and differences, both of them have potential impact on the infringement issue of AI-generated work.

---

[7] Wang X, 'A Brief Analysis of the Copyright Issues Involved in the Input and Output of Generative Artificial Intelligence - Comparison between the First Domestic AIGC Copyright Dispute Case and Foreign Cases' (IP Power, 20 February 2024) <https://www.zhichanli.com/p/1885915674> accessed 10 May 2024

[8] Peng X, ' Reproduction Rights in the Context of Technological Development and Legal Changes ' (hina Intellectual Property Research, 9 October 2007) <https://iprcn.zuel.edu.cn/zgzscqyjw-zgzscqyjw_lwxc/zgzscqyjwcn_cont_news/details-27272.html> accessed 9 May 2024

[9] 17 U.S. Code § 107 - Limitations on Exclusive Rights: Fair Use

[10] Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society[2001]OJ L 167, art 5

After the exceptions and limitations of TDM (Text and Data Mining) first introduced in EU, they has triggered objections from some rightholders who argue that they will infringe their exclusive rights of copyright. A coalition of authors, illustrators and many other artists, issued a joint statement. They said:

> Much of the copying of our works for generative AI, including "scraping" of Web pages and compilation of "datasets" for use in generative AI, has been carried out from, and/or by entities in, the European Union, claiming to rely on the exceptions to copyright for "text and data mining" (TDM) in Articles 3 and 4 of the Directive on Copyright in the Digital Single Market ("DSM Directive") enacted by the European Union in 2019.
>
> But allowing these exceptions to be applied to copying for ingestion and reuse by generative AI systems constitutes a significant violation of the obligations of EU member states as parties to the Berne Convention…
>
> We urge the European Union to promptly cure this violation of the Berne Convention and provide effective redress for the violations which have already occurred…[11]

In addition to the exceptions and limitations to copyright in these traditional legislations, as well as exceptions specifically for text and data mining, the EU's new legislation this year is also attempting to improve the relevant legal system. Including Data Act and newly adopted AI Act.

The EU's new AI Act also include provisions on the use of copyrighted content as training data, as well as references to text and data mining provisions in EU copyright law. This Act clarifies that AI training can be subject to the exceptions and limitations of the DSM Directive, and also stipulates the transparency obligations of providers. Providers need to disclose their training data, which will help rights holders know the use of their works and protect their rights.

1.2 Purpose and research questions

The purpose of this thesis is to describe and analyse the potential infringement of copyright protected works when such works are used as training data för the development of AI systems, comparing US and EU copyright law.

To fulfil the purpose set out above the following research question(s) will be answered:

1) Does the use of copyright protected works as training data constitute a reproduction under EU and US copyright law?

2) Does the use of copyright protected works as training data fall within the scope of an exception or limitation under EU and US copyright law?

3) May the generation of AI-output infringe copyright protection for copyright protected works, under EU and US copyright law?What are the provisions of EU

[11] Clark A and Calow D, 'Training AI Models: Content, Copyright and the EU and UK TDM Exceptions (via PASSLE)' (DLAPIPER, 21 December 2023) <https://mse.dlapiper.com/post/102ivrx/training-ai-models-content-copyright-and-the-eu-and-uk-tdm-exceptions> accessed 20 May 2024

and US copyright law, and what are the similarities and differences between the two?

1.3 Delimitations

This paper is limited to studying copyright infringement issues related to AI-generated artworks in Europe and the United States. The research focuses on the legal risks of AI in the creation of original artworks, and does not involve the broader application of AI technology, nor does it explore the copyrightability and copyright ownership of AI-generated works.

In addition, this study is limited to analyzing case studies and legal frameworks related to copyright law, and does not include in-depth discussions of other forms of intellectual property. In addition, this paper does not discuss in depth technical issues such as artificial intelligence, algorithms, neural networks and machine learning, but only focuses on the part related to technology and copyright law, and studies the law at the intersection of artificial intelligence and copyright law in the context of artistic creation.

1.4 Method and materials

To answer the research questions, Legal dogmatic method will be applied in this thesis. This method is a systematic, theoretical approach to legal research aimed at constructing a coherent legal system by interpreting, classifying, and systematizing existing legal texts and precedents. Through detailed analysis of legal provisions, precedents and academic literature, the internal logic and consistency of the law are determined. Focus on the interpretation and application of law, and strive to find unity and coherence within the legal system. The method will be used to study the copyright-related legal systems of the European Union and the United States.

For the copyright-related legal systems of the European Union, the method is based on the interpretation of relevant EU legal sources such as Directive (EU) 2019/790, Directive 2001/29/EC and AI Act. EU Directives and Regulations are the basis of EU judicial practice. All Member States must formulate their own domestic laws in accordance with the legal spirit of the EU level. They play an important role in guiding industry development in the EU region. Case law from CJEU (Court of Justice of the European Union) will also be used. EU case law and the Courts' interpretation of legal provisions also play an important guiding role in the trial of subsequent cases.

For the copyright-related legal systems of US. This method will be used to discuss and analyze US copyright law such as Copyright Act of 1976[12] and related cases, including those still under trial, the arguments presented in the litigation materials of both the plaintiff and the defendant, as well as the views of the court. As a

[12] See U.S. Code: Title 17. COPYRIGHTS

country of common law, judicial precedents of courts will have an important impact on subsequent case trials.[13]

Additionally, this thesis will also use comparative legal method. By comparing the similarities and differences between different law frameworks, to discuss whether AI training using protected works have risks of copyright infringement.

This thesis will also focus on some relevant and supportive literature books from copyright laws, which discuss the fair use and AI training and machine learning. The analysis and viewpoints therein provide important guidance for the discussion of this paper.

1.5 Outline

Chapter 1 is the background, purpose and research questions of this thesis. This chapter also describes the methodology.

Chapter 2 explains the technologies behind generative AI, such as machine learning and neural networks, and the connection between these technologies and copyright law, primarily regarding the right to the exclusive right of reproduction. In this chapter, it will also be discussed that the concept of reproduction rights in international law such as the Berne Convention, limitations and exceptions to copyright exclusive rights, and whether machine learning is subject to these provisions. In addition, this chapter will also discuss the copyright infringement of the output of generative AI and the scope of protection of copyright law.

Chapters 3 and 4 discuss the definition of reproduction right separately. And whether this act constitutes infringement under the legal framework of the United States and the European Union. And whether the work generated by artificial intelligence may be substantially similar to the original work because of its similarity in style.

Chapter 5 will draw conclusions.

[13] See Li Y, '[Legal Commentary] What Are the Characteristics of Anglo-American Law? How Is It Different from Civil Law?' (GAINT GROUP, 21 September 2021) <https://www.giant-group.com.tw/law-detail-1047.html> accessed 23 May 2024

2 The Technology and Copyright Law Behind Generative AI

2.1 The Technology Behind Generative AI

"Artificial intelligence or AI is used to classify machines that mimic human intelligence and human cognitive functions like problem-solving and learning. AI uses predictions and automation to optimize and solve complex tasks that humans have historically done, such as facial and speech recognition, decision-making and translation."[14] When talking about the technology behind generative AI, machine learning, deep learning and neural networks must be mentioned at the same time. These three concepts are sometimes used interchangeably. They are like a series of AI systems from largest to smallest, each encompassing the next. AI is the overarching system. Machine learning is a subset of AI. Deep learning is a subfield of machine learning, and neural networks make up the backbone of deep learning algorithms.[15]

Classic machine learning relies on human intervention. For example, if we want it to distinguish different types of fast food, human programmers need to label the salient features of these fast food for it to learn. While a deep machine learning can automatically determine the set of features to distinguish the types of fast food.[16]

Neural networks, (can be abbreviated as NNs) are sometimes called artificial neural networks (ANNs) or simulated neural networks (SNNs). A neural network is a machine learning program, or model, that makes decisions in a manner similar to the human brain, by using processes that mimic the way biological neurons work together to identify phenomena, weigh options and arrive at conclusions. One of the best-known examples of a neural network is Google's search algorithm.[17]

As for generative AI, if we want a computer to generate a specific image, it is very difficult to achieve this goal by creating a series of specific instructions. But by training an NN with images of the same category, the NN will be able to generate the required pictures autonomously.

---

[14] 'AI vs. Machine Learning vs. Deep Learning vs. Neural Networks' (IBM, 6 July 2023) <https://www.ibm.com/think/topics/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks> accessed 9 May 2024

[15] See ibid.

[16] See ibid.

[17] See 'What Is a Neural Network?' (IBM, 6 October 2021) <https://www.ibm.com/topics/neural-networks> accessed 9 May 2024

2.2 Connection between Copyright Law and AI-Generated Artworks

Data plays a crucial role for machine learning. Training requires a huge amount of data, which also brings the risk of copyright infringement. Although the current copyright laws does not contain the rightholders' exclusive right of using works for machine learning or AI training, this process may still involve the copying and storage of the works. Which are protected as exclusive rights of reproduction.

2.2.1 Definition of the Exclusive Right of Reproduction

Copyright gives protection to both moral and economic exclusive rights, the focus of this thesis is on the latter. As mentioned above, the most relevant exclusive right for machine learning technology is the right of reproduction. The definitions of "reproduction" in international law and domestic legislation of various countries are generally similar, but have differences in details between each other.

In Article 9 (1) of Berne Convention for the Protection of Literary and Artistic Works, "Authors of literary and artistic works protected by this Convention shall have the exclusive right of authorizing the reproduction of these works, in any manner or form." In this article, act of reproduction actually covers very broad with the wording of "any manner or form".[18] This statement can be fully understood to include the method of "digitizing copyrighted works", and even includes temporary reproduction behaviors such as browser caching.

In fact, the definition of reproduction rights in the Berne Convention is very contemporary. "Berne Convention was signed on 9th September 1886. The definition above was formed at the Stockholm Revision Conference in 1967".[19] At that time, areas related closely to the right of reproduction, such as traditional paper media and audio and video recordings, still rely on traditional media. The impact of digital technology on reproduction and copyright law could not be anticipated at all. This overly broad definition can no longer fully meet the current practical needs.

Article 9 of the TRIPS Agreement (Trade-Related Aspects of Intellectual Property Rights) stipulates that Article 1 to Article 21 of Berne Convention 1971 shall be complied.[20] This means that in terms of the definition of the exclusive right of reproduction, the TRIPS Agreement completely inherits the wording of Article 9 of the Berne Convention.

In Article 9 (2) of Berne Convention, it also states that countries shall "permit the reproduction of such works in certain special cases, provided that such reproduction does not conflict with a normal exploitation of the work and does not unreasonably prejudice the legitimate interests of the author."[21] This is the three-

[18] Article 9 (1) of Berne Convention for the Protection of Literary and Artistic Works.

[19] Peng X, ' Reproduction Rights in the Context of Technological Development and Legal Changes ' (hina Intellectual Property Research, 9 October 2007) <https://iprcn.zuel.edu.cn/zgzscqyjw-zgzscqyjw_lwxc/zgzscqyjwcn_cont_news/details-27272.html> accessed 9 May 2024

[20] See Agreement on Trade-Related Aspects of Intellectual Property Rights, Article 9

[21] Berne Convention for the Protection of Literary and Artistic Works, Article 9(2)

step test of Berne Convention for the exceptions and limitations of the exclusive right of reproduction.

WIPO Copyright Treaty (WCT) was adopted in Geneva on December 20, 1996. is a special agreement under the Berne Convention which deals with the protection of works and the rights of their authors in the digital environment.[22] WCT deleted the content related to the right of reproduction, including temporary reproduction.

It states that "the reproduction right, as set out in Article 9 of the Berne Convention, and the exceptions permitted thereunder, fully apply in the digital environment, in particular to the use of works in digital form. It is understood that the storage of a protected work in digital form in an electronic medium constitutes a reproduction within the meaning of Article 9 of the Berne Convention."[23]

According to this article, in the digital environment, in order to constitute the act of reproduction mentioned in WCT, the work needs to be stored. However, it is still not clearly stipulated whether this storage has a time limit. Whether the temporary storage behavior that may occur in the case of machine learning can be defined as reproduction cannot be simply determined based on this definition.

The copyright laws of the United States and the European Union also have different definitions of the exclusive right of reproduction, which will be discussed in detail below.

2.2.2 Does machine learning infringe the exclusive right of reproduction?

The copyright laws of various countries and international laws clearly list the exclusive rights protected by copyright laws. If a use of works infringes these exclusive rights without the permission of the rightholders, it will constitute copyright infringement. On the other hand, if a certain use does not infringe these exclusive rights, it does not constitute copyright infringement. At the same time, copyright laws often introduce exceptions and limitations to the exclusive rights. Thus the exclusive rights of the rightholders can be restricted. Because in the process of copyright law legislation, a balance between the protection of the exclusive rights of the rightholders and other interests, such as public interests is also crucial. In other words, use without permission does not always mean copyright infringement.

The act of using copyrighted works as training data for training AI models is not itself covered by copyright law. However, this process may still involve exclusive rights under copyright law, most likely the right of reproduction. This is because when using protected works for AI training, it is usually necessary to copy the data

[22] See WIPO Copyright Treaty

[23] Agreed statements concerning Article 1(4) of WIPO Copyright Treaty.

"in random access memory, hard disk, central processing unit, or dedicated microprocessor designed for neural network technology".[24]

Reproduction may occur at various stages of AI training, but it is usually a temporary intermediate behavior. Such reproduction may infringe the exclusive rights of rightholders in both the US and EU copyright legal systems. The specific definitions of these two and reasons will be explained in detail below.

2.2.3 Does AI Output Infringe Copyright?

There are two possible similarities between the AI-generated work and the original work: one is substantial similarity, and the other is stylistic similarity. The former is relatively simple. The generated work may directly copy the original work or part of the original work. The latter situation is more complicated and more common in practice.

To prove copyright infringement, a plaintiff must prove the the the defendant made copying of the protected elements of the work, [25] or a "striking similarity" between the two works.[26] In 1946, Arnstein v. Porter case, the US court stated two elements of establishing copyright infringement of substantial similarity: The defendant had access to the copyrighted work and the works are substantially similar.[27] The court applied two-step analysis in assessing the similarity, which are the extrinsic test and the intrinsic test.[28] The extrinsic test is whether there are overlaps of elements between the works based on objective criteria. While the intrinsic test is "whether the ordinary, reasonable person would find 'the total concept and feel of the works' to be substantially similar,"[29]

In the case of AI-generated artwork, it is not usual that the output work is very similar to the original work. In most cases, there are only similarities on style or a small number of elements. Which means it is difficult to identify as substantial similarity.

In another hand, the similarity in style does not constitute copyright infringement. The idea–expression distinction is usually adopted to analyze the scope of copyright protection. Article 9(2) of TRIPS Agreement states about the scope of copyright protection as "Copyright protection shall extend to expressions and not to ideas, procedures, methods of operation or mathematical concepts as such."30 Article 2 of Berne Convention also states that "works shall not be protected unless they have been fixed in some material form."31 Style, as an abstract concept,

---

[24] Vesala, J., & Ballardini, R. (2019). 'AI and IPR Infringement: a Case Study on Training and Using Neural Networks'. In R. M. Ballardini, P. Kuoppämäki, & O. Pitkänen (Eds.), Regulating Industrial Internet through IPR, Data Protection and Competition Law Kluwer Law International, p5-6

[25] Pasillas v. McDonald's Corp. , 927 F.2d 440, 442 (9th Cir. 1991)

[26] Baxter v. MCA, Inc. , 812 F.2d 421, 423 (9th Cir. 1987)

[27] Arnstein v. Porter, 154 F.2d 464 (2d Cir. 1946)

[28] Three Boys Music Corp. v. Bolton, 212 F.3d 477 (9th Cir. 2000)

[29] Pasillas v. McDonald's Corp, 927 F.2d at 442 (quoting Krofft , 562 F.2d at 1164)

[30] Agreement on Trade-Related Aspects of Intellectual Property Rights, Article 9(2)

[31] Berne Convention for the Protection of Literary and Artistic Works, Article 2

should be attributed to ideas rather than expressions, and cannot be described as "fixed" in Berne Convention. Therefore it cannot be protected by copyright law.

2.2.4 Conclusion

Whether the AI-generated artworks infringe copyright needs to be measured from both the output and input aspects.

From the perspective of output artworks, if the AI output works are substantially similar to the original works, then the way to determine this infringement is no different from the plagiarism of traditional works of art. However, if the generated works are only similar in style to the original works, they cannot be identified as infringements due to the protection scope of copyright law.

From the input stage of AI, that is, the stage of AI training and machine learning, it is first necessary to consider whether AI training infringes the exclusive rights of rightholders in copyright law. If it does infringe the exclusive rights, then it is necessary to consider whether such infringement can enjoy exceptions and limitations of the exclusive copyright.

In the process of machine learning, reproduction is likely to occur. However, this reproduction behavior is usually a temporary intermediate behavior. The specific definition of the exclusive right of reproduction in the copyright laws of different countries and international laws has similarities but also differences. There are differences in the details of whether reproduction includes storage behavior, whether such storage should be temporary or permanent, and whether short-term intermediate behavior is infringing.

Therefore, the reproduction behavior in the training of AI may infringe the reproduction right of the artwork, but in practice, this determination requires more technical details.

There are no specific regulations on exceptions and limitations to exclusive rights for machine learning technology. It will be discussed in the following thesis based on the Articles of copyright laws and similar cases.

3 Copyright Infringement in AI Generated Artworks under Copyright Law of US

3.1 Whether the Use of Data for AI training is Reproductions Under US Copyright Law

If the output of generative AI does not infringe copyright, what about the input? From the discussion above of machine learning, it has been concluded that machine learning may infringe the exclusive right of reproduction. What we will discuss next is whether the use of data for AI training is reproductions under US copyright law.

US copyright law protect the exclusive right of reproducing the copyrighted work in copies or phonorecords.[32] According to 17 U.S.C. § 101, the right "to reproduce the copyrighted work in copies or phonorecords" means the right to produce a material object in which the work is duplicated, transcribed, imitated, or simulated in a fixed form from which it can be "perceived, reproduced, or otherwise communicated, either directly or with the aid of a machine or device."As under the present law, a copyrighted work would be infringed by reproducing it in whole or in any substantial part, and by duplicating it exactly or by imitation or simulation.[33]

In the case of using protected works as training data, the works are likely to be reproducing in whole. However, this act of reproducing is an intermediate reproductions. Does this intermediate reproductions defined as producing unauthorized copy, different courts may give different conclusions.

According to 17 U.S.C. § 101, a copy must be "fixe in a tangible medium of expression" in a state that is "sufficiently permanent or stable to permit it to be ... reproduced...for a period of more than transitory duration."[34]

In MAI Systems Corp. v. Peak Computer, Inc.(1993), the court stated that copying and storing of software by computer repairers in the course of repair constitutes "unauthorized copying" under copyright law and therefore infringes copyright.[35] However, in Cartoon Network, LP v. CSC Holdings, Inc., this precedent was overturned. The court stated that the fixation requires a work is embodied for " more than a transitory duration."[36] But this case hadn't draw a clear boundary for the duration of "transitory". At the mean time, the MAI Systems' fixation

[32] 17 U.S. Code § 106 - Exclusive rights in copyrighted works.

[33] See Historical and Revision Notes of 17 U.S. Code § 106 - Exclusive rights in copyrighted works, house report no. 94–1476, https://www.law.cornell.edu/uscode/text/17/106

[34] 17 U.S.C. § 101 (2019).

[35] MAI Systems Corp. v. Peak Computer, Inc., 991 F.2d 511 (9th Cir. 1993)

[36] Cartoon Network LP, LLLP v. CSC Holdings, Inc., 536 F.3d 121, 129-30 (2d Cir. 2008)

standard has not been overruled and would still be followed by courts in the Ninth Circuit.[37]

Therefor, the jury is still out on whether the use of a protected work for AI training infringes the right of reproduction. But the possibility exists.

3.2 Fair Use

If the use for AI training can be defined as reproduction under copyright law, the next question is whether it is fair use. According to Copyright Act of the USA, for purposes such as criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research, is not an infringement of copyright, even though those uses has violated the exclusive rights of the original works[38]. However, this list is not a closed list, can not exhaust all potential situations of fair use. In practice, "Supreme Court of the USA has stated that fair use requires case-by-case analysis rather than bright-line rules".[39]

To analyze a certain usage could be fair use or not, there are also four factors: "(1) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes; (2) the nature of the copyrighted work; (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and (4) the effect of the use upon the potential market for or value of the copyrighted work."[40] The fact that a work is unpublished shall not itself bar a finding of fair use if such finding is made upon consideration of all the above factors.[41]

The following will analyze whether using copyrighted works to train artificial intelligence models is fair use, basing on these four factors.

3.2.1 Purpose and Character of the Use

(1) Commercial nature

The first essential element of the factor of purpose and characters of the use is commercial nature.[42]According to case laws of US courts, the commercial nature does not make it impossible for fair use.

Representative case of this element is Campbell v. Acuff-Rose Music, Inc.. Luther Campbell composed a song called "Pretty Woman". It was a rap song, a parody of which was based on Roy Orbison's rock ballad, "Oh, Pretty Woman" without a

[37] Jessica L. Gillotte, 'Copyright Infringement in AI-Generated Artworks' (2020) 53 UC Davis L Rev 2655, p2679
[38] 17 U.S. Code § 107 - Limitations on exclusive rights: Fair use
[39] Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569 (1994)
[40] 17 U.S. Code § 107 - Limitations on exclusive rights: Fair use
[41] Ibid.
[42]See ibid

license. After the about 250,000 copies of this song's recording had been sold, Acuff-Rose Music, the right holder of "Oh, Pretty Woman" , sued Campbell's band 2 Live Crew and its record company for infringing the copyright of "Oh, Pretty Woman".[43] In Campbell case, the Court of Appeals's rule runs counter to the long common law tradition of fair use adjudication about the commercial purpose. The Supreme Court cleared that commercial natural is only one element of the first factor enquiry into its purpose and character, not the decisive one.[44]

In fact, in recent years , the courts' opinions expressly minimized the importance of the commerciality inquiry.[45] That is, just because a use has a commercial purpose does not mean that the use cannot be fair. This is undoubtedly beneficial to the fair use of AI training, as the development of AI, whether directly or indirectly, often has a commercial purpose.

Therefore, concluded by referring to this case, even if the AI model may be used for commercial purposes ultimately, it cannot be judged that such use is unfair, simply because of the commercial nature.

(2) Transformative

Another essential element is if the use if transformative or not.[46]The purpose of fair use doctrine is to guarantee the breathing space within the confines of copyright.[47] If there are excessive restrictions on transformative use, it will undoubtedly hinder creation as well.

However, if the transformation is overemphasized, the economic interests of the work and the author will be affected. In the case of Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith, although disputes over the transformative nature persisted, the court ruled the infringement.[48] On this question, there is still controversy in judicial practice and academia. For AI generative works, artists who believe that they have been infringed can find the possibility to safeguard their interests from these cases.

In Blanch v. Koons case, fashion photographer Andrea Blanch sued appropriation artist Jeff Koons for copyright infringement. Koons used a picture of woman's lower legs in his own painting. This picture was taken from one of Blanch's works with our license. On Koons' work, he also added three other women's legs. During the trial of this case, a discussion about transformation was hold. "Transformation requires creating within the allegedly infringing work new or different purposes or functions as compared to those of the original."[49] The court hold that Koons used this picture to express different means of the original work. In Blanch's work, she

[43] See Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569 (1994)
[44] See ibid
[45] See Neil Weinstock Netanel, 'MAKING SENSE OF FAIR USE', (2011)
[46] Harper & Row v. Nation Enterprises, 471 U.S. 539 (1985)
[47] See Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569 (1994)
[48] See Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith,598 U.S. (2023)
[49] Jessica L. Gillotte, 'Copyright Infringement in AI-Generated Artworks' (2020) 53 UC Davis L Rev 2655

wanted to show sexuality to the photographs. Which Koons wanted the viewer to think about his/her personal experience by this objects.[50]

Similar with Koons' work, as appropriation art, in Andy Warhol Foundation case, the court came to a different conclusion. Lynn Goldsmith took a picture of Prince Rogers Nelson in 1981. In 1984, Andy Warhol created a series of silkscreen printing works based on this picture, which contained 16 works in this series. But the license he got from Goldsmith was a single time license. Andy Warhol Foundation published one of these works and received 10,000 US dollars for the copyright licensing without any payments to Goldsmith.

Referring to Blanch's case, it sees that these two cases are similar to each other. They both transformed the original work into a different form of art. Andy Warhol gain his reputation by his silkscreening works of famous people, such as the Marilyn Diptych (1962). Although these works seem meeting the requirement of transformation, the core question of this case was back to the commercial nature. "The district judge should not assume the role of art critic and seek to ascertain the intent behind or meaning of the works at issue", wrote Judge Gerard Lynch of the US Court of Appeals for the Second Circuit. "That is so both because judges are typically unsuited to make aesthetic judgments and because such perceptions are inherently subjective."[51]

Before Andy Warhol case, applying the old reasoning to the context of Al-generated artwork, the use for training of generative AI model is highly transformative. As the original purpose of an artist is to communicate through artistic expression, earn a living, and/or practice their craft. By contrast, the engineer converts such expressive works into training data to allow an Al program to refine its algorithm by minimizing the error between generated works and the ideal output, which is informed by the training data, as previously described.[52] AndyWarhol case complicates this simple analysis. In this case, both works by Andy Warhol and Goldsmith were licensed the use of magazines. Which support the option that the former work possibly make a risk of the income for the latter in its potential market. As for AI generative art works, the original artist can simply licensing their works to other AI platform to make a similar situation with the Warhol case. Which raises serious doubts about the strength of fair use as a defense to allegedly infringing creations of generative AI platforms.[53]

---

[50] See Blanch v. Koons, 467 F.3d 244, 247-49 (2d Cir. 2006)

[51] See Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith, 598 U.S. (2023)

[52] Jessica L. Gillotte, 'Copyright Infringement in AI-Generated Artworks' (2020) 53 UC Davis L Rev 2655, p2684

[53]See Eichner S, 'Warhol's Ghost in the Machine: What Goldsmith v. Warhol Means for Generative AI'(IPWatchdog, 9 June 2023) <https://ipwatchdog.com/2023/06/08/warhols-ghost-machine-goldsmith-v-warhol-means-generative-ai/id=162175/> accessed 9 May 2024

3.2.2 The Nature of the Copyrighted Work

Mostly the factor one and four are more important among the four factors of fair use. The second factor, the nature of copyrighted work, is "of the creative or instructive type that the copyright laws value and seek to foster."[54]

This thesis mainly discusses the AI-generated artworks, which are mainly paintings and illustrations, and are undoubtedly creative works.

In Blanch case, the court discussed this factor, but just in a brief way. The two focuses in this case was (1) whether the work contained more creative character or factual or informational character.[55] Because Blanch's original photograph was publicly accessible, this factor also weighed in favor of a finding of fair use.[56]

As for generative AI, the copyrighted works used by AI platforms are usually published on the internet. Therefor, this factor would possibly in favor of fair use. However, among the four factors analyzing fair use, the role this factor plays is not as essential as others. Its impact on the final conclusion about whether the AI generative work fair use or not is also limited.

3.2.3 The Amount and Substantiality of the Portion

The third factor of fair use is the amount and substantiality of the portion used in relation to the copyrighted work as a whole.[57] For this factor, the focus is whether the amount is "reasonable in relation to the purpose of the copying."[58]

In the context of an Al-generated artwork, the purpose of the use is training AI , which is completely different with the original purpose when the author creating the work. Therefor, the court is likely to find the amount of the use is fair.

3.2.4 The effect of the use upon the potential market for or value of the copyrighted work.

The fair use terms imposes the condition that fair use should not excessively damage the market for the original work by serving as a viable substitute for it.[59] When "two works usually serve different market functions," they are likely not substitutes.[60] Thus, in the context of Al generated artwork and AI training, the question is whether the AI-generated works and the original works have possibility

---

[54] HathiTrust, 755 F.3d at 96 (quoting Pierre N. Leval, Toward a FairUse Standard, 103 HARV. L. REV. 1105, 1117 (1990)).

[55] Blanch v. Koons, 467 F.3d 244, 256 (2d Cir. 2006).

[56] Jessica L. Gillotte, 'Copyright Infringement in AI-Generated Artworks' (2020) 53 UC Davis L Rev 2655, p2685

[57] See id.

[58] Blanch, 467 F.3d at 257 (citing Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 586 (1994) (quoting Folsom v. Marsh, 9 F. Cas. 342, 348 (1841)).

[59] Jessica L. Gillotte, 'Copyright Infringement in AI-Generated Artworks' (2020) 53 UC Davis L Rev 2655, p2685-2686

[60] See Campbell, 510 U.S. at 570

completing within a same potential market, or whether the AI-generated works have possibility to replace the original works.

In the field of art collection, it is reasonable to assume that AI-generated artworks will not do great harm to original artistry. For collectors, the value of an artwork does not only lie in the painting itself, but also the stories behind the painting.

A very famous example is The Next Rembrandt. It was a 2016 project run by ING Bank, J. Walter Thompson Amsterdam, and Microsoft. After analyzing all 346 of Rembrandt's paintings to identify the key common features, they made a new Rembrandt's portrait by AI and 3D printing.[61] Rembrandt was a Dutch Golden Age painter, printmaker, and draughtsman. He is generally considered one of the greatest visual artists in the history of art. He was dead in 1669.[62] In accordance with international copyright conventions or EU copyright law, his works have expired the duration of copyright protection and entered the public domain. Therefore, there is no risk of infringement when using his works to train AI or create new works imitating his style. Emmanuel Flores, director of technology for the project said "Our goal was to make a machine that works like Rembrandt… we will understand better what makes a masterpiece a masterpiece… I don't think we can substitute Rembrandt. "[63] His words made sense. In fact for artists like Rembrandt, new technology will never hinder the value of their works. Compared with the paintings themselves, collectors are more concerned about the reputation of artists and artworks, the stories behind these artworks, the connotations that the artists want to convey through the paintings, or the value of these artists and artworks in art history. Especially in the context of the market for artworks by reowned artists, consumers are unlikely to simply buy a print of the original artwork.[64]

At the same time, AI-generated artworks have entered the market as a new, independent art form. In 2016, a Turkish artist sold an Al-generated artwork entitled GCHQ for $8000 at a Google charity auction in San Francisco.[65] On October 25, 2018, for the first time in history, an Al-generated work created by the art collective Obvious was auctioned at Christie's in New York City.[66] The work, entitled Portraitof Edward Bellamy, sold for $432,500, about forty-five times higher than what Christie's initially estimated.[67] In the above two cases, the AI generated artworks will hardly pose a threat to the original work in the potential market, so it is likely to be determined by the court as fair use.

---

[61] See 'The next Rembrandt' (VML) <https://www.vml.com/work/next-rembrandt> accessed 9 May 2024

[62] See 'Rembrandt' (Wikipedia, 20 May 2024) <https://en.wikipedia.org/wiki/Rembrandt> accessed 24 May 2024

[63]See Baraniuk C, 'Computer Paints "new Rembrandt" after Old Works Analysis' (BBC News, 6 April 2016) <https://www.bbc.com/news/technology-35977315> accessed 9 May 2024

[64]See Dr. Nikki Martinez PsyDL, 'The Art of Buying Art' (HuffPost, 12 September 2017) <https://www.huffpost.com/entry/the-art-of-buying-art_b_59b3603ae4b0bef3378ce048> accessed 9 May 2024

[65] See Gaskin S, 'When Art Created by Artificial Intelligence Sells, Who Gets Paid? | Artsy' (Art Market, 17 September 2018) <https://www.artsy.net/article/artsy-editorial-art-created-artificial-intelligence-sells-paid> accessed 9 May 2024

[66] See CBS News, 'This AI-Generated Portrait Just Sold for a Stunning $432,500' (CBS News, 26 October 2018) <https://www.cbsnews.com/news/ai-generated-portrait-sells-for-stunning-432500-portrait-of-edmond-de-belamy/> accessed 9 May 2024

[67] See CHRISTIE'S, 'The First Piece of AI-Generated Art to Come to Auction' (Christie's) <https://www.christies.com/en/stories/a-collaboration-between-two-artists-one-human-one-a-machine-0cd01f4e232f4279a525a446d60d4cd1> accessed 10 May 2024

But in the field of illustrations for commercial uses, it's probably the exact opposite situation. In Andy Warhol case, both works were used for the cover of magazines, which affected the final judgment badly.[68]

The situation between digital illustrations and AI-generated works is very similar. Both of them are likely to be used in markets such as the Internet, advertising or publishing licensing. As for the use of AI training, the original authors can also authorize other AI platforms as training data, or prove that platform have the license of their training data, proving the existence of this potential market, and the economic losses that this use may cause to themselves in this potential market. With reference to the Andy Warhol case, which is likely to have a huge impact on the court's final decision of fair use.

[68] See Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith,598 U.S. (2023)

3.3 Case Analyze

Andersen v. Stability AI Ltd.

Three artists have filed a class action lawsuit against three artificial intelligence companies on behalf of themselves and other artists on January 13, 2023. Challenging the defendant's use of their copyrighted artworks in developing generative artificial intelligence. Then the defendants filed separate motions to dismiss. On October 30, 2023, the U.S. District Court issued an order, finding that their complaint was "defective in many respects" and was deficient in substantial ways. Defendants' motion granted. But they are allowed to "leave to amend to provide clarity regarding their theories of how each defendant separately violated their copyrights, removed or altered their copyright management information, or violated their rights of publicity and plausible facts in support."[69] On November 29, 2023, the plaintiff artists filed an amended complaint and added seven artists as plaintiffs, and Runway AI est. 2018 was also added as the fourth defendant.[70]

Although some cases that have similarities to AI training data have been cited above to analyze the infringement risks and the possibility of fair use in this case. There are no cases that directly target artificial intelligence training data yet. Therefore, the outcome of this case will have a significant impact on this field.

（1）Output: Substantial similarities between AI outputs and the artists' copyrighted works.

By the AI platforms of these companies, users may text descriptions of their ideal images, after a short wait, the AI will generate the images illustrating their instructions. If the descriptions contains a certain style, the generated image will also reflect this style correctly without any indication of copyright ownership. For the artists, this means they may lose many clients because of these AI platforms. Because customers no longer need to invite artists to create works, nor do they need to wait for a long creation time or pay for the license, they can get the artworks they want through AI.

When advertising and promoting these generative AI, they said "you're all gonna get [your] mind blown by this style feature …it has cores and punks and artist names … as much as we could dump in there … i should be clear it's not just genres its also artist names … it's mostly artist names … 4000 artist names."[71] Which proves that this imitation of the artist's style was not accidental but deliberate.

The plaintiff artists argued that this imitation of style constituted a substantial similarity because the AI-generated work imitated their "recurring visual elements and artistic techniques".[72] The defendant AI platform did not agree with this

[69] Andersen v. Stability AI Ltd. (3:23-cv-00201) District Court, N.D. California.

[70] Schrader A, 'In a Blow for Artists, a Federal Judge Has Sided with Three A.I. Companies in a Copyright Dispute' (Artnet News, 6 December 2023) <https://news.artnet.com/art-world/federal-judge-sides-with-ai-companies-in-artists-copyright-dispute-2387654> accessed 11 May 2024

[71] Andersen v. Stability AI Ltd, First Amended Complaint, Andersen (No. 23-cv-00201-WHO) (N.D. Cal. Nov. 29, 2023), at 72

[72] Andersen v. Stability AI Ltd, First Amended Complaint, at 10, Andersen (No. 23-cv-00201-WHO) (N.D. Cal. Nov. 29, 2023).

statement. However, copyright laws does not protect style. This means such substantively similar claims may not be recognized by the court.

（2） Input：Infringement and fair use in machine learning.

The plaintiff artists believe that "AI image products are trained on vast numbers of copyrighted images without consent, credit, or compensation and violate the rights of millions of artists."[73]This actually discusses whether it is legal to produce copies and store protected works and use these works for AI training.

The image-generating software Stable Diffusion, was developed by Stability, and be used by all the defendants.[74] This software relies on LAION (Large-Scale Artificial Intelligence Open Network), [75]LAION makes large-scale machine-learning models and datasets public, including an immense datasets of images that have been used to train machine-learning models, including Stable Diffusion.[76] This datasets contains all the plaintiffs' protected works. But LAION does not store copies of these works. It refers them by URL. This means all the defendants have to producing local copies by themselves without license.[77]Which may infringe the exclusive rights of reproduction.

However, the defendants take the opposite view. AI art tools contain NO copies of images. "Neural network weights and biases capture, to put it simply, the statistical relationships between elements – for images, things like shape, colour, position, etc – and in effect function as the reverse of image recognition: making something that's not recognizable become more recognizable."[78]

Unfortunately, this case is still under trial. We can not know the final verdict. But there is no doubt that this judgment will bring a great impact on the development of generative AI and the artwork market in the future.

3.4 Conclusion

Analyzing whether a case is fair use or not always requires a case-by-case analysis. There are still some relevant cases in the judicial process. We can not have the final conclusions yet. However, the Andy Warhol case should have some implications. Fair use is intended to guarantee breathing space within the confines of copyright.[79]

---

[73] Andersen v. Stability AI Ltd, First Amended Complaint, at 1, Andersen (No. 23-cv-00201-WHO) (N.D. Cal. Nov. 29, 2023).

[74] Order on 49, 51, 52, 58 Motions to Dismiss and Strike by Judge William H. Orrick, Andersen v. Stability AI Ltd., (No. 23-cv-00201-WHO, N.D. Cal., Oct. 30, 2023), at 2.

[75] Andersen v. Stability AI Ltd, First Amended Complaint, at 13, Andersen (No. 23-cv-00201-WHO) (N.D. Cal. Nov. 29, 2023).

[76] Andersen v. Stability AI Ltd, First Amended Complaint, at 13, Andersen (No. 23-cv-00201-WHO) (N.D. Cal. Nov. 29, 2023).

[77] Andersen v. Stability AI Ltd, First Amended Complaint, at 13, Andersen (No. 23-cv-00201-WHO) (N.D. Cal. Nov. 29, 2023).

[78] Butterick M, 'Stable Diffusion Frivolous · Because Frivolous Lawsuits Based on Ignorance Deserve a Response.''(Stable Diffusion Frivolous, 13 January 2023) <http://www.stablediffusionfrivolous.com/> accessed 8 May 2024

[79] Blanch v. Koons, 467 F.3d 244, 256 (2d Cir. 2006).

During the creation of artworks, artists always have to learn and reference from previous works. Fair use is to protect this reasonable space for their learning and reference. However, fair use is a provision of copyright law after all. The purpose of copyright protection is to encourage innovation by protecting the rights of original authors and ensuring their reasonable returns. There seems to be a contradiction between the two. In this case, instead of simply questioning whether generative AI is fair use, we should also think about which way the balance should lean.

Judging from the events caused by AI-generated artwork in recent years, especially in the field of commercial illustration, the harm is obvious. In particular, AI platforms are often enterprises with greater influence and voice than original authors. In this case, if original authors cannot receive reasonable remuneration for their creations, the meaning of copyright protection will be lost.

4 Copyright Infringement in AI-Generated Artworks under Copyright Law of EU

Compared to the United States, which still has no targeted AI regulatory bill, the EU, although lagging behind in the practice of AI technology, has taken the lead in making progress in legislation. Therefore, when discussing the infringement of AI training in the EU, in addition to discussing the situation under traditional copyright laws, the newly introduced AI regulation will also have an impact on this issue.

4.1 The Definition of Reproductions Under EU Law

In the EU legal order, according to Article 2 Information Society Directive, the definition of right of reproduction is "any direct or indirect, temporary or permanent reproduction by any means and in any form, in whole or in part" and is reserved to the right holder of copyright works and other protected subject-matter.

This definition is very broad, making it enough to "extend to every act of reproduction, however transient or irrelevant it may be from an economic perspective."[80] This means that no matter what the purpose of the reproduction is, whether it is temporary or not, it will be restricted. For example, caching when browsing the Internet and websites is also covered by this definition. Whatever the nature of the reproduction behavior produced by machine learning, it is covered under this concept Therefore, it is necessary to set exceptions and limitations for this exclusive right.

4.2 Exceptions and Limitations in EU Legal Framework

Intellectual property rights are not unlimtied. In the common-law countries, the fair use doctrine is prevalent. As the fair use of US Copyright Act discussed above in this thesis. Meanwhile, civil law countries usually adopted lists of exceptions and limitations. Three-step test in international law has the same effect, which is also adopted by some national laws. In Article 5 of InfoSoc Directive of EU, exceptions and limitations are ruled as a exhausted list of cases. There are two restrictions in EU copyright legal instruments that are relevant to generative AI and machine learning.

[80] Thomas Margoni, 'Artificial Intelligence, Machine Learning and EU Copyright Law: Who Owns AI?' (November 10, 2018), CREATe Working Paper 2018/12 (December 2018)

4.2.1 Exception of Temporary Acts of Reproduction in InfoSoc Directive

In Article 5(1) of InfoSoc Directive, temporary acts of reproduction are adopted into the list of exception and limitations. According to this Article, the temporary acts of reproduction are transient or incidental and an integral and essential part of a technological process and whose sole purpose is to enable: (a) a transmission in a network between third parties by an intermediary, or (b) a lawful use of a work or other subject-matter to be made, and which have no independent economic significance.[81]

The EUCJ offered some interpretative guidance on the conditions listed in Article 5 (1) and recital 33, including that this exception should include acts which enable browsing as well as acts of caching to take place.[82] The Court clarified that, the temporary reproduction acts need to be entirely in the context of the implementation of the technological process. It also has to be necessary for the function of the technological process. This condition is satisfied notwithstanding the fact that initiating and terminating that process involves human intervention.[83]Moreover, temporary reproduction acts must pursue a single purpose of enabling the lawful use of the protected works. Which means the use is permitted by the right holder or not restricted by law. Thirdly, these temporary reproduction acts can not produce an additional profit, or lead to a modification of that work.[84]

In the context of machine learning, temporary copies of the training data are enough for neural network. These copies do not need to be stored permanently. Therefor, from this point, using protected works for AI training meets the requirement of the exceptions of InfoSoc Directive. However, the use of copyright works in this context must satisfy a three-step test.[85]

Accoeding to Article 5(5) of InfoSoc Directive, these exceptions are only permitted (1) in certain special cases; (2) which do not result in a conflict with the normal exploitation of a work and (3) which do not unreasonably prejudice the legitimate interests of the author (or other right-holder).[86]

Therefor, the exception may not applied for machine learning in some cases. For instance, in New York Times v. OpenAI and Microsoft, Times states that the large language models “can generate output that recites Times content verbatim, closely summarizes it, and mimics its expressive style.” Which damages their relationship with readers, and make impact on “subscription, licensing, advertising, and affiliate revenue.”[87] In this case, if the plaintiff can prove their interests are indeed damaged

---

[81] Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonization of certain aspects of copyright and related rights in the information society [2001] OJ L167/10, art. 5(1).

[82] Recital 33 of ibid.

[83] Case C-302/10 Infopaq International A/S v Danske Dagblades Forening [2012]

[84] See Case C-5/08 Infopaq International A/S v Danske Dagblades Forening [2009] ECR I-6569 and C-302/10 Infopaq International A/S v Danske Dagblades Forening [2012]

[85] Ted Shapiro & Sunniva Hannson, “The DSM Copyright Directive-EU Copyright Will Indeed Never Be the Same”, EIPR 404, 6 (2019)

[86] Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society[2001]OJ L 167, art 5

[87] The New York Times Company v. Microsoft Corporation (1:23-cv-11195)

as a result of defendants' AI training, this exception may not apply for their behaviour.

4.2.2 TDM Exception in DSM Directive

TDM (text and data mining) is "a term commonly used to describe the automated processing ('machine reading') of large volumes of text and data to uncover new knowledge or insights".[88]TDM usually requires producing copies of a big amount of material, extracting the relevant data, and recombining it to identify patterns.[89]But with the widespread application of related technologies, TDM is also constantly changing and improving.

The first question is whether AI training is TDM or not. The answer should be positive. In Article 2 of DSM Directive, TDM is definite as "any automated analytical technique aimed at analyzing text and data in digital form in order to generate information which includes but is not limited to patterns, trends and correlations"[90] This definition is quite broad. It can cover the acts of machine leaning.

DSM Directive introduced two Articles of TDM exceptions, which are Article 3 and Article 4. There are differences in the scope of application of the two Articles, which will be discussed in detail below. Both exceptions are now mandatory, meaning that Member States shall introduce them into their national law, which is helpful for the harmonization of TDM policy within EU internal market.

Acts about TDM (text and data mining) for the purposes of scientific research are introduced in Article 3 of Digital Single Market Directive.[91] However, this Article sets some conditions. The first one is the nature of the entities. Research organizations and cultural heritage institutions can use this exception. According to Article 2(1) of DSM Directive, Research organizations are defined as "a university, including its libraries, a research institute or any other entity, the primary goal of which is to conduct scientific research or to carry out educational activities involving also the conduct of scientific research: (a) on a not-for-profit basis or by reinvesting all the profits in its scientific research; or (b) pursuant to a public interest mission recognized by a Member State; in such a way that the access to the results generated by such scientific research cannot be enjoyed on a preferential basis by an undertaking that exercises a decisive influence upon such

---

[88] European Commission, Commission Staff Working Document ─ Impact Assessment on the Modernization of EU Copyright Rules Accompanying the Document Proposal for a Directive of the European Parliament and of the Council on Copyright in the Digital Single Market and Proposal for a Regulation of the European Parliament and of the Council laying down Rules on the Exercise of Copyright and Related Rights Applicable to Certain Online Transmissions of Broadcasting Organizations and Retransmissions of Television and Radio Programmes, SWD (2016) 301 final, Part 1/3, §4.3.1.

[89] UK Intellectual Property Office, Supporting Document T Text Mining and Data Analytics in Call for Evidence Responses (2011) <https://webarchive.nationalarchives.gov.uk/20140603125140/http://www.ipo.gov.uk/ipreview-doc-t.pdf> accessed 22 May 2024.

[90] Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC (Text with EEA relevance) [2019] OJ L130/92. art. 2

[91] Ibid. art. 3

organisation"[92] While in Article Article 2 (3), cultural heritage institutions are defined as "a publicly accessible library or museum, an archive or a film or audio heritage institution."[93]

From these two definitions, it is obvious Article 3 is aiming to give the exceptions to the non-profit organizations or institutions. However, this Article excludes profitable AI development.

But in practice, cooperation between companies and institutions with different nature is quite common. In Recital 11 of DSM Directive, it is allowed that research organizations and cultural heritage institution "should be able to rely on their private partners for carrying out text and data mining, including by using their technological tools."[94] Which means, profitable enterprises can benefit from this Article through partnership with the research organizations and cultural heritage institution.

The second condition is the purpose for the use of TDM. To enjoy the TDM exceptions in Article 3, it has to be for the purposes of scientific research. This means although profitable AI companies may enjoy exception by cooperating with research organizations and cultural heritage institutions, the use of TDM is still limited within the scope of their cooperation, which is the research purpose. They can not create economic interests from it.

The next condition is " lawful access".[95] "By lawful access" means "access to content based on an open access policy or through contractual arrangements between rightholders and research organizations or cultural heritage institutions, such as subscriptions, or through other lawful means […] Lawful access should also cover access to content that is freely available online."[96] Which means the TDM exception of Article 3 can not be used on protected material. This will undoubtedly cause great restrictions to TDM exceptions. This Article also stipulates the obligation of research organizations and culture heritage institutions to securely store the data.

In addition, the TDM exception for research is binding for the parties and cannot be overridden by contract.[97] Article 4 of DSM Directive is the exception or limitation for text and data mining in a broader field. Besides the requirement of the "lawful access" of data is same with Article 3, this Article has a wider scope of application, covering all kinds of TDM under the definition of Article 2, including the TDM with machine learning process. This Article allows anyone to do text and data mining, without a limitation for beneficiaries or research purpose,[98] but unlike

[92] Ibid. art. 2 (1)

[93] Ibid. art. 2 (3)

[94] Ibid. Recital 11

[95] Ibid. art. 3

[96] Ibid. Recital 14

[97] Rossana Ducato and Alain M. Strowel, 'Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out' (2021) European Intellectual Property Review

[98] Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC (Text with EEA relevance) [2019] OJ L130/92. art. 4

Article 3, the right holders can restrict TDM by contract or via a machine-readable means.[99]

In the context of machine learning, there are some uncertainty and restriction of these Articles.

(1) The scope of Article 3 is narrow and unclear.

Recital 12 states that "scientific research" refers to works both in natural and human sciences. The question is, is computer science or AI development under the concept of natural or human sciences.

A comparison can be made with the concept of "research" in other EU legal documents. According to recital 159 of the General Data Protection Regulation, research is defined " in a broad manner including for example technological development and demonstration, fundamental research, applied research and privately funded research. In addition, it should take into account the Union's objective under Article 179(1) TFEU of achieving a European Research Area. Scientific research purposes should also include studies conducted in the public interest in the area of public health".[100]

If this concept is interpreted narrowly in Article 3, then computer science may not be covered by the concept of scientific research.

(2) Opt-out freely under Article 4.

As for Article 4, although it seems more general. But it leaves the right of choice to the rightholders. Rightsholders can express their choice through contracts, unilateral statements or other various means. Platforms can also prohibit TDM by their online terms. Therefor, in practice, The impact of Article 4 could be greatly limited.

However, there are also some criticisms believes that the right to opt-out is unnecessary.

The right to opt-out amounts to economically inefficient overprotection of copyright. Free use of media content for Generative AI training does not affect media sales to consumers. Opt-outs only strengthen the bargaining position of copyright holders who decide in function of their private interests. That generates windfall profits without any increase in consumer surplus or social welfare.

Free use of media content for Generative AI training does not affect media sales to consumers. Opt-outs only strengthen the bargaining position of copyright holders who decide in function of their private interests. That generates windfall profits without any increase in consumer surplus or social welfare. [101]

---

[99] Rossana Ducato and Alain M. Strowel, 'Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out' (2021) European Intellectual Property Review

[100] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data[2016] OJ L119/1. Recital 159

[101] Martens B, ' Why Should EU Copyright Protection Be Reduced to Realise the Innovation Benefits of Generative AI? ' (Bruegel, 11 March 2024) <https://www.bruegel.org/newsletter/why-should-eu-copyright-protection-be-reduced-realise-innovation-benefits-generative-ai> accessed 24 May 2024

This view has some truth. The development of AI technology relies on data, and this model of completely opt-out by the rightholders is not conducive to the acquisition and use of data by AI technology developers.

However, from another perspective, although the marginal cost of generated images with AI is very low, it is unreasonable to measure the profit margin of AI based on each generated image independently. Developers are not philanthropists. They have many profit models, and generative AI is undoubtedly an effective tool for them to make profits.

In addition, although the use of artificial intelligence training is not a normal use and profit method for artists, overall, training is not an isolated action. The ultimate goal of training is to generate images, and the images generated by it, with the advantages of low price, fast speed, and accurate imitation of the artist's style, will undoubtedly have a negative impact on the artist's regular income.[102]

Therefore, this completely free exit mechanism may be not the best choice, but it is also unreasonable to completely deprive artists of their right to choose. Perhaps a better way to solve this problem is to comprehensively weigh the training behavior and output results, and balance the legal rights of artists and the needs of technological development.

(3) Other limitations coexist

As mentioned above, in Article 5 of the InfoSoc Directive, the three-step test is applied. In Article 7(2) of DSM Directive, it states that "Article 5(5) of Directive 2001/29/EC shall apply to the exceptions and limitations provided for under this Title. The first, third and fifth subparagraphs of Article 6(4) of Directive 2001/29/EC shall apply to Articles 3 to 6 of this Directive."[103]

It can be seen from many cases that, when interpreting the three-step-test in Article 5(5) InfoSoc Directive, European Court of Justice usually interprets it restrictively. For instance, in the judgement of ACI Adam BV and Others v Stichting de Thuiskopie and Stichting Onderhandelingen Thuiskopie vergoeding (C-435/12), the court stated that "Consequently, Article 5(5) of Directive 2001/29 is not intended either to affect the substantive content of provisions falling within the scope of Article 5(2) of that directive or, inter alia, to extend the scope of the different exceptions and limitations provided for therein."[104] For this reason, this test may not be very restrictive for the application of TDM exception in practice.

In people's general impression, the business models of traditional publishing and art industries are unlikely to conflict with the areas involved in TDM technology.Which means it will not "conflict with a normal exploitation of the

[102] Handong Wu, 'Research on the Fair Use System of Copyright'(4th ed., China Renmin University Press, 2020), p131

[103] Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC (Text with EEA relevance) [2019] OJ L130/92. art. 7 (2)

[104] par.26, ACI Adam BV and Others v Stichting de Thuiskopie and Stichting Onderhandelingen Thuiskopie vergoeding (C-435/12), Judgment of the Court (Fourth Chamber) of 10 April 2014, ECLI:EU:C:2014:254.

work."[105] However, in the context of the Internet economy, many traditional media, including books, magazines, newspapers, etc., have been trying to develop more business models, not to mention areas such as commercial illustrations that are already highly dependent on the Internet. Therefore, the three-step test could have a certain impact on the application of this TDM exception.

(4) Lawful access as a precondition for TDM

Lawful access is a precondition for TDM. Users can not use TDM technology as a tool to access to the data they want.

Lawful access includes methods such as subscriptions, as well as free access to content on the Internet. For small companies or individual developers who cannot afford to pay, there is still a risk of infringement.

4.3 The Software and Database Directives

Article 6 of Database Directives adopted the exceptions to restricted. In this Article it states that " the performance by the lawful user of a database or of a copy thereof of any of the acts listed in Article 5 which is necessary for the purposes of access to the contents of the databases and normal use of the contents by the lawful user shall not require the authorization of the author of the database. Where the lawful user is authorized to use only part of the database, this provision shall apply only to that part."[106]

In Article 8 it added the rights and obligations of lawful users as: "1. The maker of a database which is made available to the public in whatever manner may not prevent a lawful user of the database from extracting and/or re-utilizing insubstantial parts of its contents, evaluated qualitatively and/or quantitatively, for any purposes whatsoever. Where the lawful user is authorized to extract and/or re-utilize only part of the database, this paragraph shall apply only to that part."[107]

"2. A lawful user of a database which is made available to the public in whatever manner may not perform acts which conflict with normal exploitation of the database or unreasonably prejudice the legitimate interests of the maker of the database."[108]

"3. A lawful user of a database which is made available to the public in any manner may not cause prejudice to the holder of a copyright or related right in respect of the works or subject matter contained in the database."[109]

---

[105] Article 5(5) of Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society [2001] OJ L167/10

[106] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data[2016] OJ L119/1. art. 6(1)

[107] Ibid. art. 8

[108] Ibid.

[109] Ibid.

In addition, Article 15 of the Directive once again clarifies that the protection of the above rights cannot be changed through contract.

TDM acts are in accordance with the normal use of this Directive and are therefore protected and cannot be restricted by contract.

4.4 AI Act

The Artificial Intelligence Act (AI Act) was proposed by the European Commission on 21 April 2021, and passed in the European Parliament on 13 March 2024. Its scope would encompass all types of AI in a broad range of sectors (exceptions include AI systems used solely for military, national security, research, and non-professional purpose).[110] Although this Act mainly aims to classify and regulate AI applications based on their risk to cause harm, rather than copyright protection. Many of its provisions will still have a significant positive impact on copyright protection related to AI-generated works.

In Article 105 of AI Act, it stated the huge demand for data in AI technology and required that all kinds of use of protected works require authorization from the rightholders, unless it meets the exceptions and limitations of Article 3 and Article 4 of DSM Directive . This means that TDM and the use of data for the purpose of training AI are not necessarily considered to meet the exceptions and limitations for TDM in the DSM Directive.

This article explicitly links the TDM exceptions and limitations of DSM Directive to AI, ending the controversy over whether the exception applies to AI training.

"The development and training of such models require access to vast amounts of text, images, videos, and other data. Text and data mining techniques may be used extensively in this context for the retrieval and analysis of such content, which may be protected by copyright and related rights. Any use of copyright protected content requires the authorisation of the rightsholder concerned unless relevant copyright exceptions and limitations apply.

Directive (EU) 2019/790 introduced exceptions and limitations allowing reproductions and extractions of works or other subject matter, for the purpose of text and data mining, under certain conditions. Under these rules, rightsholders may choose to reserve their rights over their works or other subject matter to prevent text and data mining, unless this is done for the purposes of scientific research.

[110] Council of the EU, 'Artificial Intelligence Act: Council and Parliament Strike a Deal on the First Rules for AI in the World - Consilium ' (European Council, 9 December 2023) <https://www.consilium.europa.eu/en/press/press-releases/2023/12/09/artificial-intelligence-act-council-and-parliament-strike-a-deal-on-the-first-worldwide-rules-for-ai/> accessed 22 May 2024

Where the rights to opt out has been expressly reserved in an appropriate manner, providers of general-purpose AI models need to obtain an authorisation from rightsholders if they want to carry out text and data mining over such works."[111]

One of the central aspects of the AI Act is the imposition of transparency requirements on providers.[112] In Recital 107, it mandates AI providers make the content used for AI training available for public, including both public and private sources. This will make it possible for rightsholders to exercise their rights effectively.

in addition, in Article 53 (2), it also applies the obligation of transparency information. Requiring AI providers to provide necessary information about the data they have used.

"Providers of general-purpose AI models shall:

(b) draw up, keep up-to-date and make available information and documentation to providers of AI systems who intend to integrate the general-purpose AI model into their AI systems. Without prejudice to the need to respect and protect intellectual property rights and confidential business information or trade secrets in accordance with Union and national law, the information and documentation shall:

(i) enable providers of AI systems to have a good understanding of the capabilities and limitations of the general-purpose AI model and to comply with their obligations pursuant to this Regulation; "[113]

The content that needs to be disclosed is also listed in this Act. Among them, the main ones related to the training of generative artificial intelligence are as follows:
" Transparency information referred to in Article 53(1), point (b) - technical documentation for providers of general-purpose AI models to downstream providers that integrate the model into their AI system. The information referred to in Article 53(1), point (b) shall contain at least the following:

1. A general description of the general-purpose AI model including:

...

(g) the modality (e.g., text, image) and format of inputs and outputs;

2. A description of the elements of the model and of the process for its development, including:

...

---

[111] Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts, COM/2021/206 final, Document 52021PC0206. art.105

[112] KEA, 'Eu Ai Act: Shaping Copyright Compliance in the Age of Ai Innovation' (KEA, 14 March 2024) <https://keanet.eu/eu-ai-act-shaping-copyright-compliance-in-the-age-of-ai-innovation/> accessed 22 May 2024

[113] Article 53 Obligations for providers ofProposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts, COM/2021/206 final, Document 52021PC0206

(b) the modality (e.g., text, image, etc.) and format of the inputs and outputs and their maximum size (e.g., context window length, etc.);

(c) information on the data used for training, testing and validation, where applicable, including the type and provenance of data and curation methodologies."[114]

In addition, the AI Act also has provisions on jurisdiction.

Article 106 of this Act emphasizes that regardless of whether the training of the AI model takes place within the EU or not, the AI providers must comply with relevant EU acts when placing their products on the EU market. "Any provider placing a general-purpose AI model on the Union market should comply with this obligation, regardless of the jurisdiction in which the copyright-relevant acts underpinning the training of those general-purpose AI models take place"[115]This ensures that all providers adhere to consistent copyright regulations, thereby promoting fair competition and safeguarding the interests of creators and rightsholders.[116]

4.5 Conclusion

In terms of AI commercial practices, the EU's development speed is relatively backward compared to the United States and Asia. However, in terms of legislation on AI, the EU is undoubtedly very advanced.

Overall, the EU's protection of the exclusive rights of right holders is very complete. Although there are exceptions and limitations for information network technology, including TDM, which limit some of the exclusive rights of right holders, the EU also reserves the right of right holders to choose to withdraw voluntarily. Both the DSM Directive and the AI Act emphasize that in addition to data mining for research purposes, right holders have the right to refuse companies from using their protected works through data mining in various ways. For AI providers, this means that data mining still poses a risk of infringement.

Compared with the right holder, this flow of choice means that data mining still carries the risk of infringement for AI providers. The technical principles of artificial intelligence and machine learning determine that if you want to train a more powerful artificial intelligence model, you will inevitably need a huge amount of data input. Since the right holder retains the right to choose, in order to avoid the risk of infringement, AI model providers have to pay more practice costs and licensing fees to obtain the data they need. The increase in corporate costs is undoubtedly detrimental to the development of artificial intelligence technology.

---

[114] ANNEX XII of Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts, COM/2021/206 final, Document 52021PC0206

[115] Ibid. Art. 106.

[116] KEA, 'EU AI Act: Shaping Copyright Compliance in the Age of Ai Innovation ' (KEA, 2 March 2024) <https://keanet.eu/eu-ai-act-shaping-copyright-compliance-in-the-age-of-ai-innovation/> accessed 23 May 2024

5 Conclusion

Artworks generated by artificial intelligence may infringe copyright in the input stage, that is, the training stage of artificial intelligence, and in its output works.

However, for the output works, unless the work is substantially similar to the original work, it is difficult to be identified as infringement due to the scope of protection of copyright law. Copyright law protects expression, not ideas. It is not normal for AI to simply collage or plagiarize the original work. Once it happens, the judgment method is not much different from the traditional copyright infringement judgment. So there is no need to pay too much attention to this situation.

However, in reality, there are more and more complicated situations where AI imitates the style of the original work. The generated work has no substantial similarity with the original, but can accurately imitate the artist's lighting, composition and other styles. This style belongs to the category of ideas that are not protected by copyright law. Therefore, if an artist believes that works generated by AI have harmed his or her interests by imitating his or her style, it will be quite difficult for him or her to obtain compensation through an infringement determination of the output work.

Therefore, if artists want to seek protection for their economic rights through copyright law, they must shift their focus to the artificial intelligence training stage.

However， It is still unclear whether the use of data during the training phase of artificial intelligence, that is, the machine learning phase, infringes on proprietary rights.

In the process of machine learning, it is possible to copy and store the training data. Therefore, the most relevant proprietary right for machine learning is the right of reproduction.

There is currently controversy over the possibility of infringement of the right of reproduction involved in machine learning can be subject to the exceptions and limitations to exclusive rights in copyright law.

There are some differences in the definitions of this issue in the domestic laws of different jurisdictions and in international agreements related to copyright.

According to the above, it can be concluded that whether from international intellectual property laws and agreements represented by the Berne Convention or from the copyright laws of the United States and the European Union, although there are certain differences between these laws, the intermediate reproduction behavior generated by machine learning is likely to be identified as an infringement of the exclusive right of reproduction.

Another controversy is whether such reproduction behavior can be subject to the exceptions and limitations to exclusive rights in copyright law.

International treaties such as the Berne Convention, as well as fair use in U.S. copyright law and exceptions and limitations in EU copyright law, can constrain the exclusive rights of right holders, making it possible for AI providers to use protected works without worrying about infringement liability. And whether such behavior complies with the exceptions and limitations to exclusive rights in copyright law is still controversial. Because these regulations do not explicitly address machine learning technology.

However, new progress is constantly emerging in this field. The numerous pending cases in the United States and new legislative attempts in the field of artificial intelligence in the European Union are constantly promoting the improvement of regulations in this field, increasing judicial predictability, and providing more guidance for the development of the industry.

The Andy Warhol verdict in the United States sparked a huge discussion. This case emphasizes that the four elements of fair use cannot be measured one-sidedly. Based on the spirit of the judgment in this case, artificial intelligence training may also be considered an infringement.

In the EU, the DSM Directive is still vague about machine learning technology, but the AI Act has made clearer provisions for artificial intelligence technology. It clearly states that artificial intelligence training falls under the limitations and exceptions of Articles 3 and 4 of the DSM Directive. In particular, the transparency obligations of AI providers stipulated therein require the disclosure of training information, which will greatly facilitate the supervision of artificial intelligence and the protection of rightholders.

On the other hand, in the commercial practice of the field of artificial intelligence, both artists and artificial intelligence platforms are in urgent need of clear rules. How will the law structure our era where human creative expression and growing artificial intelligence capabilities intersect. How will we continue to safeguard artists' rights, freedom of speech, and intellectual property protection?

AI image products are using artificial intelligence to generate images that are obviously similar to the plaintiff's artworks, some of which are "indistinguishable" from their own works, as artists have pointed out. These AI works imitate the artist's unique aesthetic style and have the advantages of AI's cheapness and speed. If they are completely exempted from liability for copyright infringement, it will be easy to encroach on the market and harm the rights and interests of human artists. Artists have the right to be fairly compensated for their labor like other industries.

Even more unfairly, AI companies seem to do this intentionally, advertising their services to create works in the style of specific artists. Currently, many for-profit AI company platforms provide users with such simulation services.

The artists' lawyer Matthew Butterick said, "even assuming a nominal loss of $1 per image, the value of this misappropriation would be approximately $5 billion.

(By comparison, the largest art theft in history was the theft of 13 artworks from the Isabella Stewart Gardner Museum in 1990, currently estimated to be worth $500 million.)" [117]

On the other hand, technological progress is an unstoppable trend. Intellectual property regulations will always need to strike a balance between the exclusive rights of the right holder and the broader public interest. Machine learning requires the use of a large amount of data. If copyright law strictly restricts the use of data in machine learning and requires them to obtain authorization and pay fees, it will be a huge obstacle to the development of its technology and will also be a huge blow to small and medium-sized enterprises that cannot afford the relevant fees.

In summary, a more effective approach may be to consider both input and output comprehensively. For users of artificial intelligence and artists who believe that their rights have been infringed, what they care more about is not what happened during the training of artificial intelligence or the development of machine learning technology, but whether the works generated by artificial intelligence are related to their vital interests. It is impractical to protect style through copyright law, so seeking the possibility of infringement in the process of machine learning has become a more feasible way for them to seek relief. Therefore, perhaps the input and output of artificial intelligence should be considered more comprehensively. When measuring whether the training of artificial intelligence infringes copyright, in addition to considering whether the use of data for training itself is infringing, the ultimate purpose of the training should also be taken into consideration.

[117] Southern California Chapter, 'Because AI Needs to Be Fair & Eth - i - cal for Every - one' (SCASA) <https://imagegeneratorlitigation.com/> accessed 24 May 2024

References

Official documents

European Commission

Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society(2001)

Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC (Text with EEA relevance) (2019)

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data(2016)

Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts, COM/2021/206 final, Document 52021PC0206.

US Law

U. S. Code: Title 17. COPYRIGHTS

International Agreements

Berne Convention for the Protection of Literary and Artistic Works

Agreement on Trade-Related Aspects of Intellectual Property Rights

WIPO Copyright Treaty

Literature

Monographs

Jane C. Ginsburg, 'International Copyright Law: U.S. and E.U. Perspectives: Text and Cases' (1st ed., Edward Elgar Publishing Ltd, 2015)

Handong Wu, 'Research on the Fair Use System of Copyright' (4th ed., China Renmin University Press, 2020)

Shaojun Liu, 'Rights Restriction and Digital Technology: Reform of the Fair Use System of Copyright' (1st ed., Intellectual Property Press, 2019)

Yue Xiu, 'Copyright protection research and practice'(1st ed., Intellectual Property Press, 2017)

Articles

Jessica L. Gillotte, 'Copyright Infringement in AI-Generated Artworks' (2020) 53 UC Davis L Rev 2655

Vesala, J., & Ballardini, R. (2019). 'AI and IPR Infringement: a Case Study on Training and Using Neural Networks'. In R. M. Ballardini, P. Kuoppämäki, & O. Pitkänen (Eds.), Regulating Industrial Internet through IPR, Data Protection and Competition Law Kluwer Law International

Ted Shapiro & Sunniva Hannson, "The DSM Copyright Directive-EU Copyright Will Indeed Never Be the Same", EIPR 404, 6 (2019).

'AI vs. Machine Learning vs. Deep Learning vs. Neural Networks' (IBM, 6 July 2023) <https://www.ibm.com/think/topics/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks> accessed 9 May 2024

Baraniuk C, 'Computer Paints "new Rembrandt" after Old Works Analysis' (BBC News, 6 April 2016) <https://www.bbc.com/news/technology-35977315> accessed 9 May 2024

Clark A and Calow D, 'Training AI Models: Content, Copyright and the EU and UK TDM Exceptions (via PASSLE)' (DLAPIPER, 21 December 2023) <https://mse.dlapiper.com/post/102ivrx/training-ai-models-content-copyright-and-the-eu-and-uk-tdm-exceptions> accessed 20 May 2024

Eichner S, 'Warhol's Ghost in the Machine: What Goldsmith v. Warhol Means for Generative AI ' (IPWatchdog, 9 June 2023) <https://ipwatchdog.com/2023/06/08/warhols-ghost-machine-goldsmith-v-warhol-means-generative-ai/id=162175/> accessed 9 May 2024

Game Corner, '"Artist Collective Protest against AI Drawings Occupied ArtStation Homepage, Epic Games Failed to Appease It, but Became More and More Violent " ' (United News Network, 3 January 2023) <https://tech.udn.com/tech/story/123158/6840587> accessed 23 May 2024

Li Y, '[Legal Commentary] What Are the Characteristics of Anglo-American Law? How Is It Different from Civil Law?' (GAINT GROUP, 21 September 2021) <https://www.giant-group.com.tw/law-detail-1047.html> accessed 23 May 2024

Peng X, 'Reproduction Rights in the Context of Technological Development and Legal Changes ' (hina Intellectual Property Research, 9 October 2007) <https://iprcn.zuel.edu.cn/zgzscqyjw-zgzscqyjw_lwxc/zgzscqyjwcn_cont_news/details-27272.html> accessed 9 May 2024

Wang X, ' A Brief Analysis of the Copyright Issues Involved in the Input and Output of Generative Artificial Intelligence - Comparison between the First Domestic AIGC Copyright Dispute Case and Foreign Cases' (IP Power, 20 February 2024) <https://www.zhichanli.com/p/1885915674> accessed 10 May 2024

'What Is a Neural Network?' (IBM, 6 October 2021) <https://www.ibm.com/topics/neural-networks> accessed 9 May 2024

Dr. Nikki Martinez PsyDL, 'The Art of Buying Art' (HuffPost, 12 September 2017) <https://www.huffpost.com/entry/the-art-of-buying art_b_59b3603ae4b0bef3378ce048>

Gaskin S, 'When Art Created by Artificial Intelligence Sells, Who Gets Paid? ' (Art Market, 17 September 2018) <https://www.artsy.net/article/artsy-editorial-art-created-artificial-intelligence-sells-paid> accessed 9 May 2024

CBS News, 'This AI-Generated Portrait Just Sold for a Stunning $432,500' (CBS News, 26 October 2018) <https://www.cbsnews.com/news/ai-generated-portrait-sells-for-stunning-432500-portrait-of-edmond-de-belamy/> accessed 9 May 2024

CHRISTIE'S, 'The First Piece of AI-Generated Art to Come to Auction' (Christie's) <https://www.christies.com/en/stories/a-collaboration-between-two-artists-one-human-one-a-machine-0cd01f4e232f4279a525a446d60d4cd1>

Schrader A, ‘In a Blow for Artists, a Federal Judge Has Sided with Three A.I. Companies in a Copyright Dispute ’ (Artnet News, 6 December 2023) <https://news.artnet.com/art-world/federal-judge-sides-with-ai-companies-in-artists-copyright-dispute-2387654>

Butterick M, ‘Stable Diffusion Frivolous · Because Frivolous Lawsuits Based on Ignorance Deserve a Response.’’(Stable Diffusion Frivolous, 13 January 2023) <http://www.stablediffusionfrivolous.com/> accessed 8 May 2024

Rossana Ducato and Alain M. Strowel, ‘ Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out’ (2021) European Intellectual Property Review

Eleonora Rosati (2019) Copyright as an obstacle or an enabler? A Europea perspective on text and data mining and its role in the development of AI creativity, Asia Pacific Law Review, 27:2, 198-217, DOI: 10.1080/10192557.2019.1705525

Case law

Court of Justice of the European Union

Case C-5/08 Infopaq International A/S v Danske Dagblades Forening [2009] ECR I-656

C-302/10 Infopaq International A/S v Danske Dagblades Forening [2012]

Court of US

Andersen v. Stability Ai Ltd., 3:23-Cv-00201

Pasillas v. McDonald’s Corp. , 927 F.2d 440, 442 (9th Cir. 1991)

Baxter v. MCA, Inc. , 812 F.2d 421, 423 (9th Cir. 1987)

Arnstein v. Porter, 154 F.2d 464 (2d Cir. 1946)

Three Boys Music Corp. v. Bolton, 212 F.3d 477 (9th Cir. 2000)

Pasillas v. McDonald’s Corp, 927 F.2d at 442 (quoting Krofft , 562 F.2d at 1164)

MAI Systems Corp. v. Peak Computer, Inc., 991 F.2d 511 (9th Cir. 1993)

Cartoon Network LP, LLLP v. CSC Holdings, Inc., 536 F.3d 121, 129-30 (2d Cir. 2008)

Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569 (1994)

Harper & Row v. Nation Enterprises, 471 U.S. 539 (1985)

Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith,598 U.S. (2023)

Blanch v. Koons, 467 F.3d 244, 247-49 (2d Cir. 2006)

HathiTrust, 755 F.3d at 96

Andersen v. Stability AI Ltd. (3:23-cv-00201) District Court, N.D. California