

Copyright Infringement Defenses for Generative AI Training: Comparing U.S. Fair Use, EU TDM Exceptions and the AI Act

Ruini Du*

Xianda College of Economics and Humanities, Shanghai International Studies University, Shanghai, China

**Corresponding author: Ruini Du.*

Abstract

Training generative artificial intelligence models often involves extensive reproduction of copyrighted works. Whether such training may lawfully be conducted without authorization has consequently become a major point of contention in contemporary copyright law. This article systematically examines the legal characterization of the use of copyrighted works during the training of generative AI models, focusing on two central questions: whether reproductions made during AI training fall within the scope of copyright infringement and whether existing defenses, particularly U.S. fair use and EU text and data mining (TDM) exceptions, can justify such conduct. Through a comparative analysis of the literature and recent developments including *Andy Warhol Foundation v. Goldsmith* and *Kneschke v. LAION*, this article argues that both systems face structural limitations. In the United States, the increasingly restrictive interpretation of transformative use and growing concern over market substitution weaken the certainty of fair-use defenses. In the European Union, the practical operation of the opt-out mechanism and transparency obligations remains problematic. The phenomenon of model memorization further challenges the theoretical foundation of non-expressive use. The article concludes by evaluating collective licensing, statutory licensing, taxation mechanisms, and technological safeguards as possible reform directions.

Keywords

generative artificial intelligence, copyright infringement, fair use, text and data mining, artificial intelligence act

1. Introduction

AI models like ChatGPT, Stable Diffusion, or Midjourney rely on vast collections of text, images, music, and other digital material, Much of which might be copyrighted. GPT-3, for instance, is trained using around 45 TB of text, and Stable Diffusion is trained on billions of captioned images [1]. As AI creators gather, preserve, manipulate, and examine copyrighted material, copyright remains a relevant concern [2,3]. The applicable legal question is not whether AI models may learn from extant creations but whether the copying required for such learning constitutes copyright infringement and, if so, what defenses can justify it. As a result, the copyright analysis of AI training consists of two steps. The first is whether the necessary copying falls within the scope of the reproduction right. The second is whether fair use, text and data mining exceptions, or other limitations provide a legal justification. This paper addresses the training phase, not subsequent disputes

over output. Output-related infringement is considered only insofar as it informs the legal status of training. Model memorization is addressed only to the extent that it affects claims that AI training is non-expressive use. The paper first examines reproduction and non-expressive use, then analyzes relevant aspects of U.S. fair use and EU text and data mining, reviews recent litigation, and concludes with reflections on reform.

2. Theoretical Foundations

2.1 Reproduction During Training and the Preliminary Determination of Infringement

Reproduction right is the core right protected by copyright law. In digital contexts, it extends beyond literal copies to cover temporary or long-term digital storage, transfer, and processing of content. Article 2 of the EU InfoSoc Directive defines reproduction in a very broad way, including any direct or indirect, temporary or permanent reproduction by any means [4]. Accordingly, many activities in AI model training may fall into the scope of copyright. Training typically comprises gathering, downloading, storing, cleaning, transforming, converting, and repeatedly processing material. And even where no trained model preserves entire copies of original works, reproduction occurs at the intermediate step. In traditional copyright terms, reproduction is enough to justify legal considerations. But reproduction does not constitute infringement in and of itself. Copyright does not protect facts, ideas, or information but rather expression itself. And the argument made by most AI developers is that training models involves the extraction of statistical relationships and patterns rather than the exploitation of the expressive content of works. Furthermore, copyright law features limitations and exceptions accommodating socially valuable uses. Thus, reproduction establishes the need for further legal considerations, not liability in itself.

2.2 U.S. Fair Use and Non-Expressive Use

The United States has not created a separate copyright exception for AI training. The court turned to fair use and integrated the four factors of Article 107 of the Copyright Law, including the purpose and nature of use, the nature of the original work, the number of parts used, and the impact on the potential market. Non-expressive use theory is the core argument in support of AI training. The theory distinguishes protected expressions from unprotected information, opinions and facts, and believes that artificial intelligence systems copy works not to disseminate expressive content, but to extract patterns to support subsequent content generation. The attraction of this theory lies in its consonance with the conventional principle that copyright protects expression and does not provide a right to knowledge. But the distinction is complicated by generative AI. Unlike search engines or data-processing tools, generative models can produce new text, images, and music that closely resemble the training material. Where their output closely resembles copyrighted material, it becomes difficult to affirm that training has involved only non-expressive use. The theory is therefore influential but incomplete as a justification for AI training. Under U.S. law, the legality of AI training mainly depends on the principle of open fair use, and the core lies in whether the use meets the relevant factors. The theory of non-expressive use provides legitimacy at the level of substantive protection of copyright, but its applicable limits to generative artificial intelligence, as well as the weight that commercial purposes and market effects should be given, still need to be tested by judicial practice.

2.3 The EU TDM Exceptions

The EU has taken a different path in the Digital Single Market Directive, setting statutory exceptions for text and data mining, rather than relying on flexible judicial principles. Article 3 stipulates exceptions for research institutions and cultural heritage institutions to reproduce and extract legally acquired works for scientific research purposes. Article 4 applies to commercial actors, including many AI developers, and permits copying and extraction for TDM purposes unless right holders expressly reserve their rights through a machine-readable opt-out. This TDM framework attempts to balance innovation and copyright protection by reducing the legal uncertainty of large-scale data analysis, while preserving certain control for right holders through access requirements and opt-out rights. It is therefore not a blanket exemption but a structured exception. It is still controversial whether generative AI training can be incorporated into TDM. Traditionally, TDM refers to the automatic analysis of text and data to obtain information, while generative AI training also involves parameter adjustment, model optimization, and generation ability development, which is beyond the

scope of simple data analysis. Therefore, it is necessary to clarify the relationship between AI training and TDM, and define which parts of the training process can fall into the TDM exception.

Beyond this definitional issue, the EU framework has also exposed practical difficulties in operation. For example, the definition of 'scientific research' in Article 3 is relatively narrow, and most commercial AI enterprises can only apply the general exception of Article 4. However, the opt-out mechanism of Article 4 lacks a unified technical standard for the retention of machine-readable rights, and it is difficult for rights holders to effectively monitor whether their works are used for training, which weakens the effectiveness of the mechanism. Moreover, the TDM exception mainly solves the legitimacy problem of replication and analysis in the input phase, and the generated AI also brings downstream problems, such as the output memory of training data, which cannot be solved by the TDM exception alone. The U.S. fair use doctrine and the EU TDM exception represent two different legal approaches. The former offers flexibility but limited legal certainty while the latter provides clearer statutory boundaries but faces implementation hurdles. Neither system alone is sufficient; their practical effects in the generative AI context will depend heavily on future judicial developments.

3. Comparative Analysis and Recent Developments

3.1 The United States: Fair Use under Pressure

In the jurisprudence of fair use in the United States, the principle of transformative-use has long been at the center. The case of *Campbell v. Acuff-Rose Music* established the criterion of transformative-use [5]. The U.S. Supreme Court changed this situation in the case of *Andy Warhol Foundation v. Goldsmith*. The Court emphasized that transformation requires meaningful differences in purpose and function rather than merely stylistic differences [5,6]. Because the original photo is close to the image created by Warhol in function and purpose, the Court did not adopt the fair use defense. This change poses a substantial challenge to the long-term dependence of the generative AI industry on fair use. The core value of AI training and output is to generate content that can bear the traditional functions of human works. Developers often emphasize the novelty of technology, but the Warhol case shows that novelty alone is not enough. The court will require proof of the existence of different functional purposes, and most generative AI products are difficult to meet this standard [5,6].

In Shen's view, under the fourth fair use factor, the commercial nature of generative AI and the potential impact of its output on the market of original works will have a significant impact on the defendant. If the court finds that the content generated by AI is a competitive alternative to the original work, AI companies may face substantial difficulties if they use fair use as a defense [5]. Scholars also debate whether precedents such as *Google Books* and *HathiTrust* can be extended to the field of generative AI. These precedents mainly involve information retrieval and search functions, and do not directly deal with the problem of content generation. Because the output of generative AI can serve many of the same purposes as human works, the direct analogy of the above precedents to generative AI is relatively limited [1,6].

3.2 Image-Generation Models and the Allocation of Liability

In his analysis of the image generation system, Murray emphasized that technical processes and legal responsibilities should be distinguished [7]. Data sets such as LAION usually only collect URLs and metadata, and do not store complete image files. The diffusion model also generates output through a statistical process that includes latent representation, noise injection, and iterative denoising, rather than copying and pasting existing images. This technical reality weakens the simplistic proposition that the artificial intelligence system is only a collection of fragments of copyrighted works, but it does not eliminate legal risks. According to Murray, users may be liable if they intentionally use tips intended to generate output similar to that of a particular copyrighted work [7]. Under this view, responsibility may in some cases shift from the developer to the user who dominates the generation process. Although this argument cannot solve all copyright issues, it shows that responsibility identification cannot only examine training practice, and user behavior and output similarity still play an important role.

3.3 The European Union: TDM Exceptions and the AI Act

The German LAION case in 2024 is the first copyright case involving AI training data sets in the EU. The Hamburg District Court found that LAION meets the exception conditions for scientific research TDM, so it can be protected by Article 3 [2,8]. The judgment clarifies the application of the TDM exception to some data set activities, and also exposes the existence of regulatory gaps in entities that only provide data without providing AI models. Article 53(1)(c) of the EU AI Act requires generic AI model providers to establish copyright compliance policies, in particular to identify and respect the reservation of rights under Article 4 (3) of the CDSM Directive; Article 53(1)(d) further requires the provider to make a full and detailed summary of the content used in the training and to publish it to the public [2,8]. At the same time, the EU AI Act introduces copyright-related obligations for general AI model providers. Article 53 requires the implementation of copyright compliance policies and the release of sufficiently detailed training data summaries [2,8]. These measures aim to improve transparency and promote copyright implementation. However, this obligation is mainly aimed at model providers and does not apply to upstream data providers. The transparency summary may be too general, making it difficult for creators to judge whether specific works are used. More importantly, transparency cannot solve the structural defects of the exit mechanism. According to Buick, the lack of uniform machine-readable standards and high transaction costs undermine the effectiveness of reservations [8]. Quintais pointed out a deeper institutional mismatch. The AI Act operates mainly through public-law enforcement and administrative penalties rather than private compensation mechanisms [2]. Therefore, even effective enforcement may not provide direct remedies for individual copyright holders. The interaction between copyright law and AI regulation is still incomplete.

3.4 Memorization and Technological Risks

The debate surrounding memorization reflects the limitations of existing legal theories. Research by Carlini and others demonstrates that large language models can reproduce training data under certain conditions. When the work appears repeatedly in the training dataset, or is more easily retrieved due to a unique pattern, the possibility of memory occurrence will increase [1,6]. This phenomenon brings difficulties to the non-expressive use theory, because if the output is highly similar to the original, it is not convincing to understand the training only by information extraction. As a result, Sag argues that memory weakens the fair-use defense based on information extraction [1]. Henderson and colleagues also pointed out that widely circulated works face greater memory risks [6]. At the same time, memory problems should not be over-magnified. Murray emphasizes that diffusion systems are designed to generate new content rather than reproduce existing works [7]. The existence of memory risks does not mean that all outputs constitute infringement. Instead, it demonstrates that legal analysis must account for both technological design and practical results.

4. Discussion and Integrated Analysis

4.1 Competing Theories of Legality

There are three main positions in the literature. The licensing-oriented position represented by Lucchi holds that unauthorized copying should generally be regarded as a superficial infringement, and AI developers should be licensed or compensated [3]. This position is closely aligned with the traditional copyright principles, emphasizing the creator's control over the use of the work. Sag, Lemley and Casey support the application of fair use, believing that artificial intelligence training constitutes non-expressive use or 'fair learning' and should be protected by fair use [1,6]. According to this view, AI systems learn from copyrighted works in a way similar to human learning. Copyright protects expressions rather than ideas, and extraction patterns from large data sets should not automatically require authorization. From a policy perspective, excessive restrictions on AI training may hinder technological innovation and conflict with the constitutional goal of copyright law to promote the advancement of science and applied art [1]. The third position advocates institutional reform. Quintais and Buick argue that there are structural flaws in the existing defense mechanisms, which require institutional reform [2,8]. This position does not regard all artificial intelligence training as inherent infringement, nor does it agree that fair use is broadly and automatically applied, but requires recognition of the shortcomings of existing laws and the formulation of more detailed rules. In the final analysis, the

differences among the three positions focus on how legal risks are distributed between creators and artificial intelligence companies.

4.2 Institutional Limitations

Based on the above analysis, the current system faces three core limitations. The limited bargaining power of right holders represents the first major obstacle. Individual creators often lack the bargaining power necessary to effectively negotiate with major AI companies [3,5]. Shen observes that large AI companies have substantial financial and technological resources, and many creators lack the ability to effectively negotiate licenses. This asymmetry makes personalized authorization difficult to implement in practice [5]. The technological and transaction costs associated with the opt-out mechanism are equally important. The lack of standardized machine-readable mechanisms makes it difficult for rights holders to judge whether their works are used [4,8]. According to Buick, even under the transparency requirements set out in the AI Act, right holders are still unable to determine whether their work appears in a specific data set, because the transparency summary is not sufficient to identify individual works. [8]. Third, the obligation of transparency may be difficult to provide meaningful practical safeguards. The training data summary rarely identifies individual works, limiting its execution utility [2,8]. Therefore, transparency alone cannot overcome deeper structural problems.

4.3 Reform Options

In response to these shortcomings, scholars have proposed a variety of reform programs. The first is collective licensing, which is granted by a collective management organization on behalf of a large number of creators. [5]. Shen believes that a package of licenses may provide centralized negotiation channels for AI enterprises, reduce transaction costs, and ensure that creators receive compensation [5], thus overcoming the practical difficulties of personalized licensing. The statutory licensing system advocated by Geiger and Iaia is more interventionist. The scheme does not require the creator to opt out, but allows the use of the work under the premise of compulsory compensation. Proponents argue that this approach is more in line with the reality of large-scale AI training, which guarantees compensation even if direct negotiations are difficult to carry out [2]. Senftlaiben proposes to tax artificial intelligence output and use the income to compensate creators. Quintais remains skeptical of this proposal, arguing that the program lacks a sufficiently direct link to copyright use, more like a general tax, rather than a compensation mechanism based on copyright law [2]. Technological mitigation measures have also been widely discussed, including removing duplicate content from training data sets, introducing noise during training to reduce memory, using human feedback to adjust model behavior, and applying similarity filters to model outputs [1,6]. From a comparative perspective, the existing frameworks of the U.S. and the EU have failed to provide a complete solution. Rational use is flexible, but it also brings uncertainty. Although the TDM exception provides a clear legal boundary, it faces operational difficulties. In the future, regulation may need a combination of legal and technical measures rather than relying on a single doctrine.

5. Conclusion

The use of copyright-protected works in generative AI training has become one of the defining copyright disputes of the digital age. Comparing the US fair use system, the EU TDM exception, the AI Act, and recent judicial developments reveals significant differences and shared challenges. After the Warhol case, the US framework has come under increasing pressure, since transformative use now requires greater attention to purpose and market function. Competition between AI outputs and copyrighted works further complicates the fair use defense. The EU TDM exception provides a more structured legal basis, though the opt-out mechanism and transparency obligations remain limited in practice. Model memory also raises the fundamental question of whether AI training can still be characterized as primarily non-expressive use. No reform plan yet fully reconciles the competing interests of innovation and creator protection. Collective licensing, statutory licensing, tax proposals, and technical safeguards each target specific weaknesses but also introduce new challenges. Future research should focus on the effectiveness of technical mitigation measures, the feasibility of remuneration-based licensing, and stronger international coordination. More broadly, the challenge generative AI poses to copyright law reflects the wider transformation of knowledge production and distribution in the digital economy. Whatever regulatory arrangement is adopted, its core objective should be to ensure that

copyright law continues to perform its constitutional function, incentivizing creative production while promoting broad dissemination and use of knowledge.

References

- [1] Henderson, P., et al. (2023). Foundation models and fair use. *Journal of Machine Learning Research*, 24(400), 1-79.
- [2] Quintais, J. P. (2025). Generative AI, copyright and the AI Act. *Computer Law & Security Review*, 56, 106107.
- [3] Lucchi, N. (2024). ChatGPT: a case study on copyright challenges for generative artificial intelligence systems. *European Journal of Risk Regulation*, 15(3), 602-624.
- [4] Margoni, T., & Kretschmer, M. (2022). A deeper look into the EU text and data mining exceptions: harmonisation, data ownership, and the future of technology. *GRUR international*, 71(8), 685-701.
- [5] Shen, C. (2024). Fair use, licensing, and authors' rights in the age of generative AI. *Nw. J. Tech. & Intell. Prop.*, 22, 157.
- [6] Sag, M. (2023). Copyright safety for generative AI. *Hous. L. Rev.*, 61, 295.
- [7] Murray, M. D. (2023). Generative AI art: Copyright infringement and fair use. *SMU Sci. & Tech. L. Rev.*, 26, 259.
- [8] Buick, A. (2025). Copyright and AI training data—transparency to the rescue? *Journal of Intellectual Property Law and Practice*, 20(3), 182-192.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.

Copyrights

Copyright for this article is retained by the author (s), with first publication rights granted to the journal. This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).