

# US Copyright Office Releases “Pre-Publication Version” of Report on Copyright Issues in Generative AI Training

## Client Alerts

May 13, 2025

By: Steve Englund, Zach Marino

On May 9, 2025, the US Copyright Office released a “pre-publication version” of Part 3 of its report on Copyright and Artificial Intelligence (the Report).<sup>[1]</sup> This much-anticipated Report focuses on use of copyrighted works in the development of generative AI systems, and particularly (1) AI training-related activities that implicate copyrights, (2) the circumstances in which AI training may or may not qualify as fair use, and (3) licensing of copyrighted works for AI training. This is the culmination of a series of reports stemming from a policy study commenced almost two years ago concerning copyright and related issues raised by the widespread availability and use of AI.<sup>[2]</sup>

After setting the table with a detailed overview of the technology of AI model development, the Report’s legal analysis begins with a discussion of various AI training-related activities that can involve reproducing copies of copyrighted works. The Report explains that copyrights are implicated by producing a training dataset containing copyrighted works, various copies made in the training process (including model weights that embody substantial protectable expression from copyrighted works), reproductions made in retrieval-augmented generation, and output material that replicates or closely resembles copyrighted works.

The main focus of the Report is the fair use doctrine, which permits certain uses of copyrighted works without authorization based on a balancing of four statutory factors (described below). The Office carefully addresses application of each of the fair use factors to a range of AI training-related activities. Notably, the Report explains that, while “[v]arious uses of copyrighted works in AI training are likely to be transformative,” “[t]he extent to which they are fair . . . will depend on what works were used, from what source, for what purpose, and with what controls on the outputs—all of which can affect the market.”<sup>[3]</sup> For example, uses ultimately directed toward “analysis or research” are more likely to be fair use than “making commercial use of vast troves of copyrighted works to produce expressive content that competes with them in existing markets, especially where this is accomplished through illegal access.”<sup>[4]</sup>

The Report concludes with a discussion of licensing, finding that for now, licensing markets should be allowed to continue to develop without government intervention.

The Report states that it is being released in pre-publication form “in response to congressional inquiries and expressions of interest from stakeholders.”<sup>[5]</sup> It anticipates a “final version” to be “published in the near future, without any substantive changes expected in the analysis or conclusions.”<sup>[6]</sup> However, the fate of the Report has become less clear with the removal of the Register of Copyrights by the Administration the day after the Report was released.

## **Prima Facie Infringement**

Developing an AI system generally involves a range of copying activity. In some phases of the development process, it is clear when copies have been made.<sup>[7]</sup> However, it has been less clear when an AI model itself may be said to contain a reproduction, or be a derivative work, of material on which the model is trained. The Report explains that AI model weights can contain copies of works in a model’s training data when the model has “memorized” that data.<sup>[8]</sup> Specifically, the Report notes that, “[l]ike other digital files that encode or compress content . . . the content need not be directly perceivable to constitute a copy,” so long as it is sufficiently “fixed” and can be “perceived, reproduced, or otherwise communicated . . . with the aid of a machine or device.”<sup>[9]</sup>

## **Fair Use**

Fair use was the most contentious issue addressed in the approximately 10,000 comments the Office received as part of its AI study. The Report emphasizes the Supreme Court’s repeated admonitions that fair use analysis requires a fact-specific and context-dependent balancing of the statutory fair use factors in light of the purpose of copyright law.<sup>[10]</sup>

### **First Factor: Purpose and Character of the Use**

Analysis of the first fair use factor must focus on the “use” of a work that is alleged to be infringing.<sup>[11]</sup> The Report explains that different acts of copying require separate consideration, but “[f]air use must also be evaluated in the context of the overall use” as “training alone is rarely the ultimate purpose.”<sup>[12]</sup>

A key question under the first factor is whether, and to what degree, the challenged use is transformative. In general, the more transformative a use is, the less significant the other factors become, although “not every transformative use is a fair one.”<sup>[13]</sup> There is a variety of AI systems, so whether a particular system has a transformative purpose, and how much weight such a purpose should be given, will depend on the circumstances.

The Report explains that “training a generative AI foundation model on a large and diverse dataset will often be transformative.”<sup>[14]</sup> This is because “[t]he process converts a massive collection of

training examples into a statistical model that can generate a wide range of outputs across a diverse array of new situations.”<sup>[15]</sup> Further, the purpose of many AI models is to perform tasks different from the purpose of the works on which they are trained.<sup>[16]</sup> At one end of the spectrum, the Report explains that model training is “most transformative” when the purpose is “to deploy it for research, or in a closed system that constrains it to a non-substitutive task.”<sup>[17]</sup> For example, a model trained “on a large collection of data, including social media posts, articles, and books, for deployment in systems used for content moderation does not have the same educational purpose as those papers and books.”<sup>[18]</sup> At the other end of the spectrum, a model developed “to generate outputs that are substantially similar to copyrighted works in the dataset” (e.g., a model “trained on images from a popular animated series and deployed to generate images of characters from that series”) is not transformative.<sup>[19]</sup>

The Report recognizes that many uses will fall somewhere between the two ends of that spectrum. The Report concludes that “[w]here a model is trained on specific types of works in order to produce content that shares the purpose of appealing to a particular audience, that use is, at best, modestly transformative.”<sup>[20]</sup> The Report explains that effective restrictions on the outputs of a model can affect the analysis by making an AI system less capable of fulfilling the purpose of the works on which a model was trained, making their use in training more transformative.<sup>[21]</sup>

The Report also addresses the “commerciality” prong of the first factor analysis, noting once again that the analysis must focus on the use involved, and whether it “serves commercial or nonprofit purposes.”<sup>[22]</sup>

Finally, the Report concludes that the knowing use of a dataset that includes “pirated” works (*i.e.*, ones obtained unlawfully) “should weigh against fair use without being determinative.”<sup>[23]</sup>

### **Second Factor: Nature of the Copyrighted Work**

The Report only briefly analyzes the second factor, noting that the analysis will be fact-specific.<sup>[24]</sup>

### **Third Factor: Amount and Substantiality of the Portion Used**

The Report notes that use of whole works in AI training generally would weigh against fair use,<sup>[25]</sup> and focuses on two considerations that might lead to a different result.

First, “[c]opying an entire work may weigh less heavily against a finding of fair use . . . where it is reasonable in relation to a transformative purpose.”<sup>[26]</sup> The Report concludes that use of whole works may be necessary for some types of training of many generative AI models and so could be reasonable when there is a transformative purpose.<sup>[27]</sup>

Second, some courts considering cases of non-public intermediate copying have focused on the portions of works ultimately made available to the public and whether they are “a competing

substitute” for the original.<sup>[28]</sup> The Report concludes that the third factor may “weigh less heavily against generative AI training where there are effective limits on the trained model’s ability to output protected material from works in the training data.”<sup>[29]</sup>

#### **Fourth Factor: Effect of the Use upon the Potential Market for or Value of the Copyrighted Work**

The fourth factor is sometimes considered the most important.<sup>[30]</sup> The Report considers various potential market effects of use of copyrighted works in AI training.

First, the Report examines potential losses in sales, finding that AI training can lead to lost sales when training uses “pirated” copies acquired from unauthorized sources, training enables a model to output (and users can readily access) substantially similar copies of the works used in training, works were specifically developed for AI training (*e.g.*, specialized training datasets), and results of retrieval-augmented generation contain protectable expression (including summaries and abridgments).<sup>[31]</sup>

Second, the Report considers “market dilution,” a form of potential market harm resulting from large numbers of AI-generated works competing with the copyrighted works used in a model’s training by saturating the market, making it more difficult for audiences to find the original works or diluting royalty pools. While acknowledging that the concept of market dilution is “uncharted territory,” the Report opines that “[t]he speed and scale at which AI systems generate content pose a serious risk of diluting markets for works of the same kind as in their training data.”<sup>[32]</sup> The Report also recognizes that outputs created to imitate a creator’s style could affect the market for their works even if those outputs are not substantially similar to the creator’s work (*i.e.*, are not infringing).<sup>[33]</sup>

Third, the Report addresses lost licensing opportunities. The Report finds a “reasonable” or “likely to be developed” licensing market for “certain copyright sectors, types of training or uses, and models,” while noting that is unclear whether there will be a sufficient licensing market “for all kinds of works at the scale required for all kinds of models.”<sup>[34]</sup>

Finally, the Report considers whether potential public benefits should affect the analysis. Finding “strong claims to public benefits on both sides,” the Report does not find copyright-related benefits from unlicensed use of copyrighted works in AI training that would affect the fair use analysis.<sup>[35]</sup>

In the end, the Report concludes that “[t]he copying involved in AI training threatens significant potential harm to the market for or value of copyrighted works.”<sup>[36]</sup>

#### **Weighing the Factors**

Courts will need to evaluate and weigh the fair use factors on a case-by-case basis. Because the uses and impacts will vary, the Report concludes “that some uses of copyrighted works for generative AI training will qualify as fair use, and some will not.”<sup>[37]</sup>

## Licensing for AI Training

The Report also summarizes comments concerning the benefits and challenges of options for licensing works for AI training. Voluntary licensing is “increasingly taking place”.<sup>[38]</sup> However, commenters identified logistical, financial, and other challenges in obtaining licenses for the number and variety of works potentially needed.<sup>[39]</sup> The comments received by the Office reflected little support for statutory licensing.<sup>[40]</sup> Ultimately, the Report “recommends allowing the licensing market to continue to develop without government intervention,” and only if it should prove necessary, considering “targeted intervention.”<sup>[41]</sup>

## Footnotes

[1] U.S. Copyright Office, Copyright and Artificial Intelligence Part 3: Generative AI Training Pre-Publication Version (May 2025), available at <https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-3-Generative-AI-Training-Report-Pre-Publication-Version.pdf>.

[2] See U.S. Copyright Office, Copyright and Artificial Intelligence Part 1: Digital Replicas (July 2024), available at <https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-1-Digital-Replicas-Report.pdf>; U.S. Copyright Office, Copyright and Artificial Intelligence Part 2: Copyrightability (January 2025), available at <https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-2-Copyrightability-Report.pdf>.

[3] Report at 107.

[4] *Id.*

[5] Report at i.

[6] *Id.*

[7] Report at 27.

[8] Report at 28–29.

[9] *Id.*; 17 U.S.C. § 101.

[10] Report at 32.

[11] Report at 36 (citing *Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith*, 598 U.S. 508, 533 (2023)).

[12] Report at 36–37.

[13] Report at 39.

[14] Report at 45.

[15] *Id.*

[16] *Id.*

[17] Report at 46.

[18] *Id.*

[19] *Id.*

[20] *Id.*

[21] Report at 46–47.

[22] Report at 50–51.

[23] Report at 52.

[24] Report at 54.

[25] Report at 55.

[26] *Id.*

[27] Report at 57.

[28] Report at 57–58 (quoting *Google Books*, 804 F.3d 202, 222 (2d Cir. 2015)).

[29] Report at 59.

[30] Report at 61.

[31] Report at 63.

[32] Report at 65.

[33] Report at 65–66.

[34] Report at 70.

[35] Report at 73.

[36] *Id.*

[37] Report at 74.

[38] Report at 85.

[39] Report at 86–87.

[40] Report at 95.

[41] Report at 106.

## Related Attorneys



### **Steve Englund**

Partner

senglund@jenner.com

+1 202 639 6006



### **Zach Marino**

Associate

zmarino@jenner.com

+1 212 407 1749

## Related Capabilities

AI Task Force

Content, Media, and Entertainment

Copyright

© 2026 Jenner & Block LLP. Attorney Advertising. Jenner & Block LLP is an Illinois Limited Liability Partnership including professional corporations. This publication, presentation, or event is not intended to provide legal advice but to provide information on legal matters and/or firm news of interest to our clients and colleagues. Readers or attendees should seek specific legal advice before taking any action with respect to matters mentioned in this publication or at this event. The attorney responsible for this communication is Brent E. Kidwell, Jenner & Block LLP, 353 N. Clark Street, Chicago, IL 60654-3456. Prior results do not guarantee a similar outcome. Jenner & Block London LLP, an affiliate of Jenner & Block LLP, is a limited liability partnership established under the laws of the State of Delaware, USA and is authorised and regulated by the Solicitors Regulation Authority with SRA number

615729. Information regarding the data we collect and the rights you have over your data can be found in our Privacy Notice. For further inquiries, please contact [dataprotection@jenner.com](mailto:dataprotection@jenner.com).

**Stay Informed**

