wiley

ALERT

# Copyright Office Issues Key Guidance on Fair Use in Generative AI Training

May 16, 2025

On May 9, 2025, the U.S. Copyright Office (the Office) released the third and final report in its "Copyright and Artificial Intelligence" series, offering its most comprehensive guidance to date on one of the most contested legal questions in the AI era: whether and to what extent the use of copyrighted works to train generative AI (GenAI) models constitutes "fair use" under U.S. copyright law. The 108-page report, *Copyright and Artificial Intelligence: Part III – Generative AI Training* (the Report), addresses growing concerns from creators, platforms, and developers about the boundaries of lawful AI development and signals a cautious but consequential interpretation of copyright's fair use doctrine in the context of GenAI.

The Report largely avoids providing a blanket endorsement or a firm rejection of fair use for GenAI training, but instead embraces nuance, recognizing that each use case is context-specific and requires a thorough evaluation of the four factors outlined in Section 107 of the Copyright Act:

> *(i) The purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes;*
>
> *(ii) The nature of the copyrighted work;*
>
> *(iii) The amount and substantiality of the portion used in relation to the copyrighted work as a whole; and*
>
> *(iv) The effect of the use upon the potential market for or value of the copyrighted work.**[1]***

## Authors


David E. Weslow
Partner
202.719.7525
dweslow@wiley.law

Scott Nuzum
Special Counsel
202.719.4953
snuzum@wiley.law

Stephanie Rigizadeh
Associate
202.719.4736
srigizadeh@wiley.law


## Practice Areas

All Things AI
Artificial Intelligence (AI)
Copyright
Intellectual Property
Intellectual Property

The Report provides a thorough technical and legal overview and takes a measured approach responding to the legal issues underlying fair use in GenAI. This alert provides an overview of the key takeaways from the Report and highlights potential implications for stakeholders stemming from the Office's analysis.

## Key Takeaways

### 1. "Transformativeness" Must Be Meaningful, Not Mechanical

In evaluating the purpose and character of the use of copyright protected content, the Office notes that courts have typically focused on the elements of "transformativeness" (i.e., use that is additive and new, of a different purpose, or not meant to substitute the original work), and commerciality (i.e., use for a commercial versus nonprofit/educational purpose).

On the issue of transformativeness, the Report concludes that GenAI training run on large, diverse datasets "will often be transformative."[2] However, the Office also affirmatively states that use of copyright protected materials for AI model training is alone insufficient to justify fair use. Instead, "transformativeness is a matter of degree"[3]; the extent to which something is transformative ultimately "depend[s] on the functionality of the model and how it is deployed."[4]

The Office notes that training a model is most transformative where "the purpose is to deploy it for research, or in a closed system that constrains it to a non-substitutive task."[5] This is contrasted with instances where the AI output closely tracks the creative intent of the input (e.g., generating art, music, or writing in a similar style or substance to the original source materials). In these instances, the Office would likely consider such usage derivative rather than transformative.

### 2. The Activity, and Not the Entity Type, Determines Commerciality

With regard to the issue of commerciality, the Office notes that a GenAI model is often the product of efforts undertaken by distinct and multiple actors, some of which are commercial entities and some of which are not.[6] It is difficult, therefore, to discern attribution and to definitively say that a model on its face is the product of a commercial or a noncommercial actor.

Even then, the Report states that just because an entity is for-profit does not mean the use will be considered "commercial" in the fair use assessment; for example, researchers within the commercial entity may well develop a model for purposes of publishing an academic research paper.[7] Likewise, a nonprofit could very well develop a GenAI model to license for commercial purposes. Accordingly, one must look beyond mere provenance of the model and the business entity structure in assessing the commerciality element. The focus of the inquiry should be on whether a case-specific use ultimately inures financial benefits and commercial purposes for the entity using copyrighted material.[8]

### 3. Use of Entire Works Can Undermine a Fair Use Defense, Especially When Made Public

While the Office acknowledges that machine learning processes often require ingestion of entire works,[9] it cautions that the wholesale taking of entire works "ordinarily weighs against fair use."[10] The critical assessment in evaluating the use of entire works in GenAI models comes down to two questions: (i) is there a transformative purpose; and (ii) how much of the work is made publicly available. Fair use is much more likely in instances where a model "entirely obscure[s] outputs from users or result[s] in non-expressive outputs."[11] Thus, where a GenAI model employs methods to prevent infringing outputs, the use of entire works for training the model may be less likely to weigh against a fair use finding.

**4. Market Harm Is a Central Concern**

In assessing market harm, the Office acknowledges that the debate of fair use in GenAI training places them in "uncharted territory."[12] According to the Report, the assessment of market harm must be analyzed more broadly, with special attention given to broad market "effects" and not merely to market harm for a specific copyrighted work.[13] The reason for this stems from the potential for AI-generated outputs to displace, dilute, and erode the markets for copyrighted works, meaning that "fewer human-authored" works are likely to be sold.[14] The Office highlights concerns raised by artists, musicians, authors, and publishers about declining demand for original works as AI-generated imitations proliferate. Where GenAI systems compete with or diminish licensing opportunities for original human creators – especially in fields such as illustration, voice acting, or journalism – the fourth factor is likely to weigh strongly against fair use.

**5. The Office Encourages Licensing Frameworks and Legislative Monitoring**

While stopping short of endorsing legislative change, the Copyright Office emphasizes the need for further development of licensing solutions. The Report calls for scalable mechanisms – whether private or collective – for obtaining rights to use copyrighted works in AI training, especially where fair use is uncertain. However, the Report also declines to endorse a compulsory licensing regime, arguing that the potential harm outweighs the benefits.[15] The Report notes the "relatively nascent"[16] state of the law, technology, and markets, and suggests that "new model architectures and techniques may be developed to facilitate training using fewer unlicensed works without sacrificing quality."

**Implications for Stakeholders**

**Developers and Technology Companies:** AI companies, especially those developing GenAI systems for text, image, music, or video generation, should proceed cautiously when incorporating copyrighted material into training datasets. The Office's analysis casts doubt on assumptions that current training practices are broadly protected under fair use. Developers should consider taking proactive steps such as licensing the content used to train their models. In addition, companies should closely monitor evolving case law (including high-profile litigation that is now pending) and be prepared to adjust business models in response to judicial or legislative developments.

**Content Creators and Rights Holders:** The Report reinforces the position of creators and rights holders who have raised alarm over the use of their works in training GenAI without permission or compensation. In cases where a GenAI model is trained on works that were pirated or illegally accessed (e.g., via circumventing

paywalls), the Office suggests that should "weigh against fair use without being determinative."[17] The Office's recognition of potential market harm, the limitations of transformativeness, and the implications of whole-work copying provides a favorable foundation for those seeking to assert control over how their works are used. Creators should explore registration, monitoring, and enforcement strategies, and consider engaging with licensing collectives or platforms that aim to facilitate permissions for GenAI training uses.

**Legal and Compliance Teams:** In-house counsel and compliance officers should treat GenAI training as a distinct area of copyright risk, separate from traditional product development or content deployment. Legal teams should assess whether their organization has sufficient visibility into the provenance of training data, the nature of any third-party datasets that are used, and the intended use of outputs. A well-structured rights clearance process coupled with indemnification provisions, particularly for commercial deployments, may be necessary to reduce litigation exposure.

**Policymakers and Industry Groups:** While the Report recommends against government intervention (for now), it anticipates further congressional interest. Policymakers will be under pressure to balance innovation with protection for creators, and industry groups should expect continued dialogue on licensing standards, metadata requirements, and transparency obligations. Voluntary codes of conduct, public-private data licensing registries, or even statutory compulsory licensing regimes may be on the horizon if private market solutions fall short.

## Looking Ahead

The Report represents a major step forward in clarifying the complex interplay between copyright law and GenAI development. While it does not provide categorical answers, it frames the debate around principled application of existing law, while encouraging industry-led solutions. Companies developing or relying on GenAI tools should reevaluate their data sourcing and risk mitigation strategies in light of this Report, and creators should be prepared to assert their rights in what is likely to continue to be a hotly contested legal battleground.

Wiley has a deep bench of attorneys with expertise in copyright and artificial intelligence issues across its multidisciplinary practice groups, including Corporate, Intellectual Property, and Telecom, Media & Technology. If you have any questions about the Copyright Office's report, please contact one of the attorneys listed on this alert.

[1] 17 USC 107

[2] Report at 45

[3] *Id.* at 46.

[4] *Id.*

[5] *Id.*

[6] *See id.*at 50.

[7] *See id.*

[8] *See id.*

[9] *See id.* at 57

[10] *Id.* at 55.

[11] *Id.* at 59.

[12] *Id.* at 65.

[13] *See id.* at 65.

[14] *See id.*

[15] *See id.* at 104.

[16] *Id.* at 105.

[17] *Id.* at *52.*