## The Copyright Office Weighs in on Use of Copyrighted Material in Generative AI Training

**May 12, 2025 | (Time to read: 3 minutes)**

**By Jenifer Wallis**

A number of industries focused on AI, most notably the legal, entertainment, and tech industries, have been waiting with bated breath on the US Copyright Office to issue guidance on one very important question: is the use of copyrighted material in generative artificial intelligence training fair use? Well, the US Copyright Office finally answered the question (kind of – the report was released pre-publication), and the answer is sometimes (maybe – the conclusion is that infringement at all stages is possible and requires a case-by-case analysis). The most noteworthy, albeit obvious, finding in Part 3: Generative AI Training is that the legal guardrails of copyright protection **do** apply in generative AI training. In other words, there is no blanket fair use defense over all instances of AI training and output. Finally, the Copyright Office recommends that the free market of licensing copyrighted material for such training continue to develop on its own, without government intervention.

On May 9, 2025, the Copyright Office released a **pre-publication** version of the highly anticipated and long-awaited Copyright and Artificial Intelligence, Part 3: Generative AI Training. The title page of the text contains this disclaimer: "The Office is releasing this pre-publication version of Part 3 in response to congressional inquiries and expressions of interest from stakeholders. A final version will be published in the near future, **without any substantive changes expected in the analysis or conclusions**." (Emphasis added).

In its 108-page report, the Copyright Office gives a comprehensive analysis of the technical background of machine learning, generative language models, and the training data and phases of such generative artificial intelligence training models. The report also provides a comprehensive discussion of prima facie copyright infringement and how such infringement could take place in all phases of generative AI training, from data collection and curation, to training, to RAG, and finally, to outputs.

Essentially, the Copyright Office found that at each step required to create and deploy a generative AI system using copyrighted material, the possibility of infringement exists. The Copyright Office also analyzed the possibility of licensing copyrighted material for use in generative AI training, noting the two opposed views on licensing: on the one hand, that such unlicensed and unauthorized use of copyrights causes immeasurable harm to creators and on the other hand, requiring AI companies to license works in training data would stifle development of the technology.

Much of the Report (52 pages)[1] is dedicated to an analysis of the factors of the fair use defense and whether it would apply to protect the generative AI use of copyrighted material. The Copyright Office found that at each stage and each factor, separate considerations must be weighed in determining whether fair use applies **and most importantly, that "[f]air use must also be evaluated in the context of the overall use**."[2] For example, an AI language model, used to "help learn a foreign language by chatting with users on diverse topics and offering corrective feedback," serves a different function than the purpose of the copyrighted works it is trained on and therefore may be transformative, lending towards fair use.[3] On the other hand, the Copyright office noted that where a model is trained "to generate outputs that are substantially similar to copyrighted works" (such as generating images of characters from popular animated series – and, logically, similar sounding songs or art derived from existing copyrighted works) "it is hard to see the use as transformative."[4]

Unfortunately for stakeholders, nothing is for certain in the current landscape. Part 3 was released as a pre-publication version, and although it states on its face that no substantive changes are expected in its analysis or conclusions, the day

after its release, the Copyright Office director was fired. Therefore, we have a non-final report that may very well undergo substantive changes with the appointment of a new Copyright Office head. Even if final, the report essentially concludes that the current legal framework of copyright case law can be used on a case-by-case basis to answer questions of infringement. Crucially, as to the issue of licensing copyrighted material for training data, "the Office recommends allowing the licensing market to continue to develop without government intervention."[5]

---

[1] Pgs. 32-84.

[2] Pg. 37.

[3] Pg. 45.

[4] Pg. 46.

[5] Pg. 106.

## Related People

- Jenifer C. Wallis