



the copyright office report on ai and fair use: a generative controversy

James D. Berkley
MSK Client Alert

May 20, 2025

Amidst a level of intrigue rare to the Library of Congress and U.S. Copyright Office, on Friday, May 9, the Copyright Office released a detailed **108-page report** containing its most extended discussion of how copyright law applies to the so-called “training” of generative artificial intelligence (better known as “generative AI”). Generative AI is the rapidly developing field exemplified by products such as ChatGPT, where large statistical models—using computer processes opaque even to their developers—can essentially “learn” by imbibing massive quantities of digitized information and then, through continued practice and fine-tuning, can become adept at generating seemingly intelligent, even seemingly creative, responses to human queries, even to the point of creating new works.

At present, more than 40 copyright lawsuits are pending across the United States involving generative AI, typically pitting owners and creators of copyrighted content against the tech company purveyors of generative AI. These cases pose complex questions over whether infringement occurs if copyrighted works are used to “train” generative AI models, and whether copyright owners are entitled to demand licenses and compensation for such uses.

Titled “**Copyright and Artificial Intelligence, Part 3: Generative AI Training**,” the Copyright Office’s report was released as an unusual “pre-publication version” on a day between two newsworthy firings. First, on May 8, the Librarian of Congress, of which the U.S. Copyright Office is part, was dismissed after serving in her role for nearly eight years. Then, the day after the report’s release, the Register of Copyrights—under whose imprimatur the report was prepared—was dismissed as well. These events prompted speculation as to whether the positions taken in the report were viewed as insufficiently pro-AI.

attorneys

James D. Berkley
Robert H. Rotstein

practice areas

Artificial Intelligence (AI)
intellectual property
specialties



the copyright office report on ai and fair use: a generative controversy

That said, the conclusions of the report do not blatantly favor either the pro-copyright or anti-AI camps. As the Introduction states, "The public interest requires striking an effective balance, allowing technological innovation to flourish while maintaining a thriving creative community."

On his website, *Copyright Lately*, MSK intellectual property partner Aaron Moss offered his "**top five takeaways**" from the report. These takeaways are summarized below:

1. Generative AI Can Implicate Different Kinds of "Copying"

As the report details, generative AI models are trained using large "datasets" that can involve making multiple copies of copyrighted works. Works must be digitized, formatted, transferred, and combined, and a completed dataset may be reproduced many times over. To no one's surprise, the Copyright Office report concludes that "**the steps required to produce a training dataset containing copyrighted works clearly implicate the right of reproduction.**" Unless these steps are defensible as "fair use," such copying may constitute infringement.

But as the report notes, downstream steps raise the issue of copying as well. If a model is trained on a set of works, and then creates output that strongly resembles one or more of those works, the output could be found to violate copyright too. A more subtle question is whether the trained model *itself* may be found infringing. The report considers whether the model's internal "weights"—the parameters that store learned information—can embody copyrighted expression. According to the Office, if a model can produce outputs that are substantially similar to the training inputs, it has memorized protectable content, and copying or distributing these weights could therefore amount to infringement. This is significant, since it would mean that even apart from the model's inputs and outputs, downloading or transmitting the trained model could be actionable copyright infringement in its own right.

2. Copying for AI Training May Be "Transformative," But Context and Purpose Matters

Copying works for purposes of training AI might still be fair use. Section 107 of the U.S. Copyright Act provides a four-factor test to consider if use of a copyrighted work is "fair." The first factor looks at "the purpose and character of the use, including whether [the] use is of a commercial nature or is for nonprofit educational purposes."

A line of Supreme Court cases instructs that even if a use is commercial, a key issue under this factor is whether the use is "transformative," meaning that it "has a further purpose or different character" from the original. And in recent cases—such as its important 2023 decision involving the magazine-cover use of an Andy Warhol portrait based on the work of a professional photographer, *Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith*, 598 U.S. 508 (2023)—the Court has instructed that courts look very carefully at the specific facts surrounding the use in question and its purpose.

The Copyright Office report follows suit, noting that for generative AI training, "**if fair use must . . . be evaluated in the context of the overall use.**" At one end of a spectrum, copyrighted works could be used in training for purposes having nothing to do with creating new works as output, let alone ones that reproduce the training data. At the other end, AI training might be used to generate works that, in copyright parlance, are "substantially



the copyright office report on ai and fair use: a generative controversy

similar" to the original works without any noticeably different purpose, which is not transformative and does not support fair use. The report adds that another factor is whether AI systems can deploy "guardrails" that prevent output from copying protected expression. If they can, using copyrighted works for training may be more transformative, and more likely to qualify as fair use. Finally, the report offers that the **source** of copyrighted works can also play a role under the first factor, stating that "the knowing use of a dataset that consists of pirated or illegally accessed works should weigh against fair use without being determinative."

3. AI Training Involves Expression, But Is Different from Human Learning

The report responds to two specific arguments sometimes raised to suggest that AI training inherently has a different "purpose" and "character," or that it should escape analysis under copyright law entirely. These arguments appeal to two widely different ideas: one, that AI training is technical and "non-expressive," the other, that AI training is akin to human learning.

The report pushes back against both of these arguments. On the first, it responds that generative AI training is more than technical, since models "absorb[] not just the meaning and parts of speech of words, but how they are selected and arranged at the sentence, paragraph, and document level," which is "the essence of linguistic expression." On the second, the report replies critiques the analogy to human learning. A human learner is not truly copying, since "humans retain only imperfect impressions of the works they have experienced, filtered through their own unique personalities, histories, memories, and worldviews." AI training, however, "involves the creation of perfect copies with the ability to analyze works nearly instantaneously," resulting in models "that that can create at superhuman speed and scale." The Copyright Office thus takes the view that AI has a fundamentally different relationship to copyright than human learning.

4. Copying Entire Works Weighs Against Fair Use, But Is Not Automatically Disqualifying

The third fair use factor under section 107 of the Copyright Act looks at "the amount and substantiality of the portion used in relation to the copyrighted work as a whole." A strike against AI training is that, generally, the entirety of the copyrighted work is used, perhaps with millions of others. But the report notes that full-work copying can still be fair use where the purpose is highly transformative, as in cases where using entire works enables valuable search tools, such as Google Books. The report finds that using entire works for AI training is "less clearly justified" than in such cases, but holds out that it might still be shown that full-scale copying is "functionally necessary" for AI models to perform optimally. Copying full works for AI training may thus be "reasonable" when there is a transformative purpose, especially if there are measures in place to avoid making copied material available as output.

5. Generative AI Poses Potential for "Market Harm"—Including Possibilities of "Flooding the Market"

The Copyright Act's fourth fair-use factor looks at "the effect of the use upon the potential market for or value of the copyrighted work." The Copyright Office's report devotes substantial space to assessing whether use of copyrighted works for AI training can pose economic harm to copyright owners. The report identifies three possible types of such harm: (1) lost sales, if AI systems are capable of creating works that substitute for the works they are trained on, (2) lost licensing opportunities, if copyright owners are deprived of the ability to be



the copyright office report on ai and fair use: a generative controversy

compensated when their works are used, and (3) market dilution, if a surge of AI-created works were to "saturate" —or "flood"—the marketplace for expressive works of a similar style or type as the training works. The third type could result in AI-based works broadly competing with human-created ones, depressing the earning potential of copyright owners even without copying or substituting for their works directly.

The Copyright Office report acknowledges this last possibility journeys into "uncharted territory." Yet the report endorses taking this theory of harm seriously, stating 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." As an example, it offers that "[i]f thousands of AI-generated romance novels are put on the market, fewer of the human-authored romance novels that the AI was trained on are likely to be sold." This position will likely be controversial, and goes beyond what courts have typically seen as the focus of fourth factor.

The above only summarizes key issues tackled in the Copyright Office's report, which also treats a range of other subjects, from AI public policy to practical considerations like licensing arrangements and collective bargaining.

What Happens Next

The current report contains much that is favorable to copyright owners, while stressing the need to be mindful of the facts and purposes that could make a particular use fair. Courts and scholars will pay close attention to the report, which notably does not carry the force of law. It is possible that the report will be rescinded, before or after a new Register of Copyrights is appointed. A new Register may bring views distinctively more pro-AI, and more inclined to favor broad, across-the-board fair-use treatment for AI training. Stay tuned.

IP Client Alert Editor: Robert H. Rotstein