

How AI Cos. Can Cope With Shifting Copyright Landscape

WRITTEN BY

Michael D. Hobbs, Jr. | Justin W. Tilghman

Published in [Law360](#) on May 28, 2024. © Copyright 2024, Portfolio Media, Inc., publisher of Law360. Reprinted here with permission.

“I would really like to take a moment to recognize all the print journalists in this room. Your words speak truth to power, your words bring light to the darkness, and most importantly, your words train the AI programs that will soon replace you.”

—Colin Jost, 2024 White House Correspondents’ Dinner

In the ever-evolving landscape of artificial intelligence development, utilizing copyrighted material for training algorithms has not only become the comedy fodder for “Saturday Night Live” cast members, but also a focal point of recent legal scrutiny.

Even a quick glance at media headlines confirms that there have been a flurry of new lawsuits and rulings in copyright infringement lawsuits against generative AI companies.

On Feb. 13, Reuters reported that a “federal judge in California has dismissed parts of a copyright lawsuit brought by comedian Sarah Silverman, Michael Chabon, Ta-Nehisi Coates and other authors against OpenAI over its alleged use of their books to train the large language model underlying its popular chatbot ChatGPT.”

On March 11, Computerworld reported that “three authors, Abdi Nazemian, Brian Keene, and Stewart O’Nan, are part of a new copyright infringement lawsuit against Nvidia, the latest such suit to challenge generative AI providers’ reliance on the ‘fair use’ doctrine to acquire copyrighted material to train their large language models.”

On March 25, Reuters reported that “Bloomberg LP has asked a New York federal judge to dismiss a lawsuit from Arkansas governor Mike Huckabee and other authors who claimed the company misused their books to train its large language model BloombergGPT.”

Reacting to the concerns about the risks from these lawsuits as well as the U.S. Supreme Court decision in the *Andy Warhol Foundation for the Visual Arts Inc. v. Goldsmith* ruling last year, many generative AI players — including Google LLC, Microsoft Corp., Amazon.com Inc., Open AI, IBM, Adobe Inc. and others — have recently announced indemnification of varying degrees for their models.

Some have offered it proactively, and others retroactively.

With courts and AI providers starting to provide answers for AI users — namely all of us — this article delves into the complex legal implications surrounding using copyrighted material in AI training and examines recent legal decisions, fair use implications and possible licensing solutions for AI users being offered by AI generators.

By way of background, at the heart of AI development lies the necessity for robust training datasets. These datasets often include copyrighted material such as images, text, and audio, and video from different sources, including books, films, databases, and the internet.

However, the materials used to train these large language models are not actual words formulated into sentences, but rather mathematical representations of patterns predicting based on statistical calculations and what comes next.

While using such material is essential for creating AI systems capable of understanding and processing real-world data, it raises significant copyright concerns that have resulted in a recent surge in copyright litigation related to AI training datasets.

One prominent case — *Kadrey v. Meta Platforms Inc.* — involves the comedian Sarah Silverman and two other writers who sued Meta and OpenAI in the U.S. District Court for the Northern District of California for using their copyrighted material to train large language models. Silverman, like many other plaintiffs, likely believed this would be an open-and-shut case of copyright infringement.

However, in November, the judge partially dismissed many of the claims. Even if refiled, the defendant's pleadings in the litigation raised complex questions regarding whether AI developers are copying content at all.

As proponents for AI developers have argued, there are no copies of the copyrighted material, but instead, the AI is using an algorithm to take the features of a work, which leads to an output that is wholly distinct and original from the copyrighted material. The multiple cases filed earlier this year claiming similar copyright infringement from the use of copyrighted content to train generative AI are sure to face similar hurdles.

More so, if there is no clear evidence of copying in the output, it will be difficult for courts to reach a finding of infringement. This puts plaintiffs in a precarious position, as they may be hesitant to fully disclose their work and creative process to show copying, balancing the need to protect their intellectual property without revealing the inspiration behind their work.

Thus, not only will plaintiffs face difficulty in litigating these types of cases, but courts will inevitably face issues applying the traditional notions of copyright law to this new technology.

Even if, however, courts can reach a threshold decision that use of copyrighted content in the training of generative AI constitutes copyright infringement, final outcomes in the cases are bound to vary, as the defense of fair use is such a fact-intensive inquiry, it makes the future legality of the creation and use of generative AI uncertain.

The doctrine of fair use provides a legal framework for the limited use of copyrighted material without the need for permission from or payment to the copyright holder.

Determining what constitutes fair use in the context of AI training was made an even more onerous task for courts in light of the Supreme Court's recent decision in *Andy Warhol Foundation v. Goldsmith*.^[1] In a 7-2 decision, the court held that the series of paintings created by artist Andy Warhol of music star Prince based on photographs of the singer by Lynn Goldsmith did not constitute fair use because Warhol's changes were not transformative.

Historically, courts in cases such as *Authors Guild v. Google* in 2015,^[2] where the U.S. Court of Appeals for the Second Circuit found that transformation of printed copyrighted books into an online searchable database through scanning and digitization was a fair use, and *Kelly v. Arriba Soft Corp.* in 2003,^[3] where the U.S. Court of Appeals for the Ninth Circuit found transformative fair use with respect to searchable images of copyrighted visual artwork, considered a work transformative when it altered the purpose of the original work.

But *Warhol* only addresses the transformation of the Warhol painting, limiting its inquiry into whether the respective works had the same commercial purpose.

The court articulated that because Goldsmith and Warhol both licensed their respective works to *Vanity Fair*, Warhol's painting was not fair use, regardless of what new message Warhol attempted to convey with his portraits of Prince.

Seemingly trying to avoid making artistic judgment calls, the court's analysis in *Warhol* centers around something more concrete — the commercial purpose of the respective works. If they serve the same commercial purpose, the use is not transformative and the defendant loses the fair use inquiry.

In the context of AI training, *Warhol* created serious litigation risks for AI developers.

While they may have been able to argue that their use of copyrighted material is transformative because they are using the material to train datasets, meaning their use was transforming the purpose of the content and thus a fair use in the same manner as *Authors Guild v. Google*, the advent of licensing agreements between copyright owners and some AI developers means that there is an existing market to license copyright content to AI developers.

Because there is a market for their content for this purpose, copyright content owners can argue, similar to Warhol's painting performing the exact same purpose of Lynn Goldsmith's photograph, namely, to provide a cover for a magazine, that the unauthorized use of copyrighted material to train AI datasets is not transformative under the *Warhol* case precedent.

While no court has yet to address this issue, the *Warhol* case has created uncertainty regarding the permissibility of using copyrighted material to train AI datasets that just a few years ago seemed in their favor.

AI developers are now at a crossroads. On the one hand they face protracted litigation from copyright content owners and a significant increase in risk of copyright infringement and a rejection of the fair use defense following *Warhol*.

They may eventually prevail, but in the meantime, they also face the crossroads of corporate and institutional clients reluctant to adopt the use of generative AI in the workplace given current litigation.

As explained in a *Forbes* article by Arun Shastri in January: “Many companies see challenges in using generative AI because they are worried about the risks. Risks come in many forms — consistency, security and infringement. The last threat, infringement, is on the minds of many leaders.”^[4]

Trying to resolve these issues, many AI developers including Google, Microsoft, Amazon, OpenAI, IBM, Adobe and others are now turning to licensing agreements with content providers to obtain the rights to use copyrighted material.

As *Reuters* reported at the end of April,^[5] *The Financial Times* had signed a deal with OpenAI to license its content for the development of AI models and allow ChatGPT to answer queries with summaries attributable to the newspaper, and conditions under which the material can be used.

This latest license relationship follows on previous deals that OpenAI has signed over the last several months with publishers Axel Springer SE, PRISA Media from Spain and French-based Le Monde Group.

In turn, having eliminated the threat of claims from copyright content owners used to train their generative AI, these developers are then contractually offering to indemnify and hold harmless users from any claims of copyright infringement.

The promise to indemnify users is a huge selling point in a market where companies are looking to utilize the efficiencies promised by AI, but avoid the risk of being swept into costly litigation such as the *Silverman* case.

For example, Microsoft updated its website in January to explain:^[6]

To address customer concern, Microsoft is announcing our new Copilot Copyright Commitment. As customers ask whether they can use Microsoft’s Copilot services and the output they generate without worrying about copyright claims, we are providing a straightforward answer: yes, you can, and if you are challenged on copyright grounds, we will assume responsibility for the potential legal risks involved.

So is this where the law on generative AI ends? That will not be the case if an existential battle breaks out similar to the one fought before the Ninth Circuit in *A&M Records Inc. v. Napster Inc.* in 2001,^[7] or the U.S. Supreme Court in *MGM Studios Inc. v. Grokster Ltd.* in 2005,^[8] where other technologies that had outpaced the law had their final reckoning with the law.

Instead, the current uncertainty of the law of generative AI ends in a series of well-negotiated licenses with copyright content owners. For the moment, the lawsuits with *Silverman* and others remain active.

But with OpenAI completing a new deal in February that values the company at \$80 billion, the future of legal disputes in this space may well be headed in the direction of well-funded settlements and agreements between content providers and the generative AI companies who will replace them.

- [1] *Andy Warhol Foundation v. Goldsmith*, (598 U.S., 2023).
- [2] *Authors Guild v. Google*, 804 F.3d 202 (2d Cir. 2015).
- [3] *Kelly v. Arriba Soft Corp.*, 336 F.3d 811 (9th Cir. 2003).
- [4] Mitigating Legal Risks When Using Generative AI, *Forbes*, January 29, 2024.
- [5] OpenAI to use FT content for training AI models in latest media tie-up. *Reuters*, April 30, 2024.
- [6] <https://blogs.microsoft.com/on-the-issues/2023/09/07/copilot-copyright-commitment-ai-legal-concerns/>.
- [7] *A&M Records, Inc. v. Napster Inc.*, 239 F.3d 1004 (9th Cir. 2001).
- [8] *MGM Studios, Inc. v. Grokster, Ltd.*, 545 U.S. 913 (2005).

RELATED INDUSTRIES + PRACTICES

- Artificial Intelligence
- Intellectual Property
- Trademark + Copyright