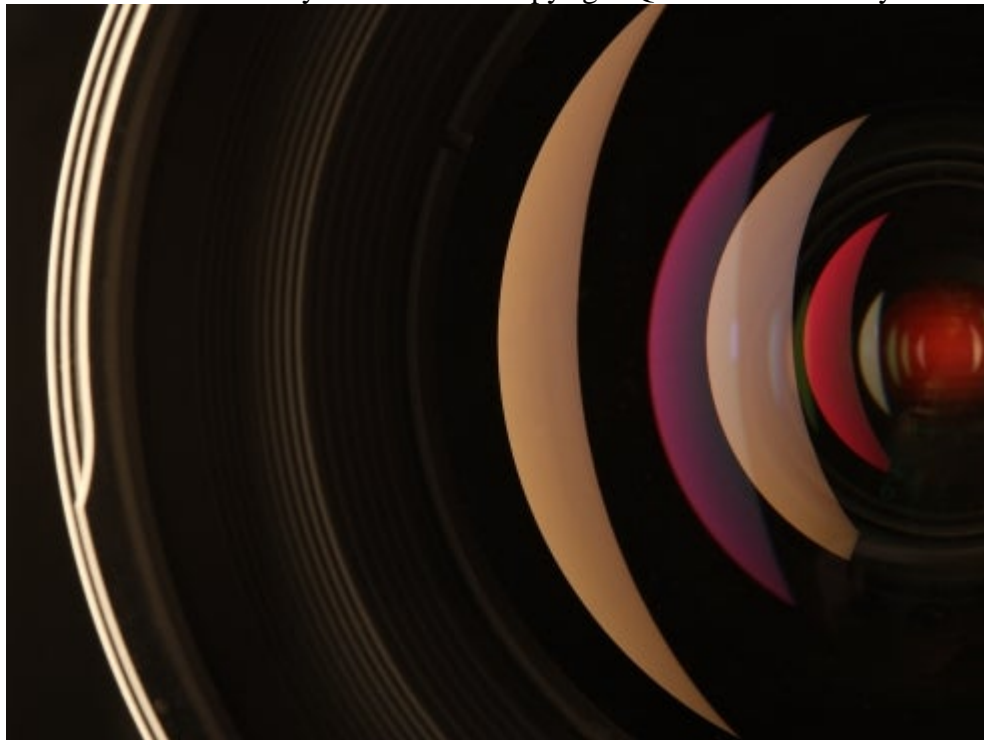


[Blogs](#)

October 25, 2023

Known Unknowns: Key Unanswered Copyright Questions Raised by Generative AI



The generative AI revolution has arrived. Will copyright law snuff it out?

Despite all the excitement surrounding generative AI tools, a cloud darkens the horizon. These tools need to be trained on massive amounts of ingested content and, according to [press reports](#), this content is often scraped without authorization from third-party websites, raising significant copyright law issues.

Indeed, a number of copyright infringement actions [have already been filed against AI tool developers](#). In January, for example, three artists who claim their works were scraped from the Internet and used to train Stable Diffusion, Midjourney, and DreamUp filed a class action against the three image generation tools. The lawsuit alleges that a generative AI tool is "merely a complex collage tool" and that every output generated by such a tool is an unauthorized derivative work of the materials used to train the tool.

The past few months have seen even more lawsuits filed against generative AI developers. Many of these recent cases have been brought by novelists, targeting copyright claims related to text generation tools and open-sourced models. Some of these cases, however, have paired copyright claims with broader allegations that AI tool providers are, to quote one complaint, "[putting the world at peril with untested and volatile AI Products](#)."

U.S. courts have yet to rule on many of the key copyright issues raised by generative AI technologies, so it's impossible to predict how these cases will play out. However, as we enter the AI era, we can identify the critical copyright law questions courts will need to address in the coming years:

Question #1: To the extent that generative AI tools are trained on third-party content, are copies of the content being made that would implicate the reproduction right of the owners of the copyrights in such content?

Under U.S. copyright law, certain "ephemeral" copies of copyrightable works, even if unauthorized by the applicable copyright owners, may *not* violate such owners' exclusive right to control reproductions of such works. For example, in [*Cartoon Network LP, LLLP v. CSC Holdings, Inc.*](#) the U.S. Court of Appeals for the Second Circuit held that a buffer copy of a TV program was not a reproduction of such program for purposes of copyright law where the buffer at issue was constantly overwritten so that, at any given moment, no more than 1.2 seconds of the program existed on the defendant's servers. Further, the court left open the possibility that a buffer copy of longer than 1.2 seconds in duration may still not run afoul of a copyright owner's reproduction rights.

Of course, whether machine learning-related copying to train a generative AI model is sufficiently ephemeral so as *not* to constitute a reproduction under the U.S. Copyright Act will depend on the techniques being used, which presumably will vary across the AI industry. Moreover, the ephemeral use exception to copyright infringement will differ in scope depending on where a dispute is litigated; for example, although, as noted above, a copy that is only 1.2 seconds in duration may not constitute copyright infringement in the Second Circuit, other circuits have yet to adopt this approach.

Question #2: Is the large-scale scraping of third-party content to train generative AI tools a fair use?

Scraping content to train generative AI tools requires creating copies of the scraped content. Assuming such copies are unauthorized and not sufficiently ephemeral to avoid infringement (see our first question above), then, absent a defense, they may infringe the applicable copyright owners' reproduction right. Fair use may provide a defense as it permits certain unauthorized uses of copyrighted works when the use is considered sufficiently beneficial to society. The Copyright Act does not define fair use, but instead sets out four nonexclusive factors for courts to consider in determining whether an unauthorized use of another's work is a fair use. The fair use factors are notoriously difficult to apply, with even experienced copyright lawyers differing on whether, in a particular instance, the factors favor the copyright owner or the alleged infringer; moreover, fair use outcomes are highly fact specific and nuanced.

While the traditional boundaries of the fair use defense had been expanding, especially where a defendant's use of the plaintiff's copyrighted work (even a commercial use) could be characterized as "transformative," the U.S. Supreme Court's recent decision in [*Andy Warhol Foundation for the Visual Arts v. Goldsmith*](#) may represent a curtailing of that expansion.

The *Goldsmith* decision, which evaluated a challenged *use* of a new work and not the new work itself, also suggests that the different phases of the training process and different applications of trained models may require separate fair use analyses, each of which may potentially have different fair use outcomes.

Question #3: Does the output of a generative AI tool constitute a derivative work of the inputted content used to train such AI tool?

Under U.S. law, in addition to their right to control reproductions of their works, copyright owners have exclusive control over the creation of derivative works based on their works. The Copyright Act broadly defines a derivative work as any work "based upon one or more preexisting works . . . in which a work may be recast, transformed, or adapted." Courts, however, typically limit that broad definition by requiring a showing that an alleged derivative work has appropriated copyrightable expression from an earlier work.

Recent complaints filed against generative AI tool providers have sought to test these traditional limits by alleging that models extracted "expressive information" from the model's training data and that model outputs are "synthesized entirely from expressive information found in the training data." The phrase "expressive information" is not used in the Copyright Act and appears intended to attempt to expand the category of works

recognized as derivative works by blurring the distinction between information, which is not protected by copyright, and creative expression, which is protected by copyright.

Even without these recent efforts, however, the line between fact and expression is not always clearly demarcated. Factual summaries of TV episodes, movies, and novels, when sufficiently detailed, have been found to constitute derivative works. Similarly, a defendant's manual containing the answers to physics problems in a plaintiff's textbook has been found to be a derivative work of the textbook.

Question #4: If the output of a generative AI tool does infringe a third-party copyrighted work, who is the direct infringer?

Although copyright infringement is a strict liability offense, U.S. courts have required "some element of volition or causation" to be established before imposing liability for direct copyright infringement. This means that a defendant must, in some meaningful way, be shown to be responsible for the creation of an infringing work before direct liability will be imposed. The volition requirement can shield the owner of a copy shop from direct liability when a customer uses the shop's photocopiers to make unauthorized copies of a work.

Ongoing splits exist both among and within U.S. circuit courts as to the meaning of "volitional conduct." The Supreme Court recently declined to hear the appeal of the Second Circuit's decision in [*ABKCO Music, Inc. v. Sagan*](#) that found direct liability only attaches to "the person who actually presses the button" to create an infringing copy. Accordingly, the circuit split on this issue will persist for the time being.

Questions of volitional conduct are further complicated by the fact that a model's generated outputs are typically influenced, but not fully controlled, by both the party that trained the model and the party that supplied the prompt and other parameters used by the model to generate an output. Additionally, models may be fine-tuned or distributed by a party other than the original model trainer.

Question #5: Could a generative AI tool provider be secondarily liable for infringing output created by tool users?

Rights holders have sought to hold AI tool providers secondarily liable for works created by the tool's users under contributory, vicarious, and inducement theories of liability. There can be no secondary liability, however, without a direct infringer. Accordingly, fair use and other potential defenses to direct infringement claims will also be relevant to questions of secondary liability.

If a direct infringement is found to have occurred, secondary liability will depend on application of the secondary liability test applicable to the theories of liability being alleged. These tests evaluate whether a defendant had knowledge of, contributed to, supervised, received a financial benefit from, or promoted the infringing activity, and test outcomes will be highly fact specific and nuanced.

Question #6: Given that the output of a generative AI tool may not be protected under U.S. copyright law, how can AI tool users protect such content?

While the U.S. Copyright Office has recognized that prompts supplied to an AI tool by a user may influence the outputs generated by the tool, the Copyright Office's current position is that the process of generating outputs is not sufficiently "controlled" by the user to allow the user to be recognized as the author of the generated work for copyright purposes.

Even if not protected by copyright, contractual restrictions may still impose controls on how generated outputs may be used. Such contractual restrictions, however, may not always be enforceable. Contractual terms that extend copyright-like restrictions to noncopyrightable materials may be preempted under the Copyright Act.

There is currently a circuit split on when contractual provisions should be held preempted, and the Supreme Court recently denied cert to a case presenting that question.

Question #7: What happens if courts find the large-scale scraping of third-party content in order to train generative AI to constitute copyright infringement?

Even if courts were to ultimately reject ephemeral and fair use arguments raised by AI companies, AI technology is simply too important to go away. There are already companies making available generative AI tools that were trained using self-developed or licensed content. Moreover, one would imagine Congress getting involved to strike an appropriate balance between the rights of content creators and the societal benefits conferred by an important new technology; Congress has done so in the past in connection with the [Section 512 DMCA safe harbors](#) and the various compulsory licenses available under U.S. copyright law that have their roots in earlier disruptive technologies such as cable television ([Section 111](#)) and even player pianos ([Section 115](#)).

We will have to wait for the courts to resolve these contested, cutting-edge issues, but we already know that how these issues are ultimately resolved will dramatically shape the future of both AI and human artistry. Despite these legal uncertainties, new generative AI tools continue to be developed and released and those tools are being widely adopted for industry and personal uses. Developers and users should, however, remain mindful of these questions when evaluating the risk these tools present, developing risk mitigation plans, and making decisions about particular use cases.

Finally, we end with the most important "known unknowns" of them all. Our seven "known unknowns" above all focus on the near future; however, stepping back and assessing the big picture, we are confronted with this overarching question as generative AI tools dramatically improve and become ubiquitous over the coming years: In a world where most content will be generated by AI, what role, if any, will copyright play? At the turn of the 19th century, U.S. cities had extensive laws and regulations specific to horses and horse-drawn carriages; automotive technology, however, eventually rendered most of these laws and regulations irrelevant. As human-created works are inevitably supplanted over time with machine-created works, will copyright law suffer the same fate? While some fear [that copyright law could destroy the generative AI revolution](#), might it be the other way around—that it is generative AI that ultimately poses an existential threat to copyright law?

Follow us on social media @PerkinsCoieLLP, and if you have any questions or comments, contact us [here](#). We invite you to learn more about our [Digital Media & Entertainment, Gaming & Sports industry group](#) and check out our podcast: [*Innovation Unlocked: The Future of Entertainment*](#).

Authors

Explore more in

[Technology Transactions & Privacy Law](#)