Generative Artificial Intelligence and Copyright Law

Updated July 18, 2025

Innovations in artificial intelligence (AI) have raised several new questions in the field of copyright law. Generative AI programs—such as Open AI's DALL-E and ChatGPT programs, Stability AI's Stable Diffusion program, and Midjourney's self-titled program—are able to generate new images, texts, and other content (or "outputs") in response to a user's textual or other prompts. Generative AI programs are trained to create such outputs partly by exposing them to large quantities of existing writings, photos, paintings, or other works.

This Legal Sidebar explores questions that courts and the U.S. Copyright Office have confronted regarding whether generative AI outputs may be copyrighted as well as whether training and using generative AI programs may infringe copyrights in other works. Other CRS Legal Sidebars explore questions AI raises in the intellectual property fields of patents and the right of publicity.

Copyright in Works Created with Generative AI

Do Copyrighted Works Require a Human Author?

The question of whether copyright protection may be afforded to AI outputs—such as images created by Midjourney or texts created by ChatGPT—hinges largely on the legal concept of "authorship." Article I, Section 8, Clause 8 of the U.S. Constitution, often referred to as the Intellectual Property (IP) Clause, empowers Congress to "secur[e] for limited Times to Authors . . . the exclusive Right to their . . . Writings." Based on this authority, the Copyright Act affords copyright protection to "original works of authorship." While the Constitution and Copyright Act do not explicitly define who (or what) may be an "author," U.S. courts to date have not recognized copyright in works that lack a human author—including works created autonomously by AI systems.

Before the proliferation of generative AI, U.S. courts did not extend copyright protection to various nonhuman authors, holding that a monkey who took photos of himself lacked standing to sue under the Copyright Act; that human authorship was required to copyright a book purportedly inspired by celestial beings; and that a living garden could not be copyrighted. The U.S. Copyright Office has also long maintained that copyrighted works must be "created by a human being" and therefore refused to register

works that are "produced by a machine or mere mechanical process that operates randomly or automatically without any creative input or intervention from a human author."

At least one lawsuit has—unsuccessfully, thus far—challenged the human-authorship requirement in the context of AI. In June 2022, Stephen Thaler sued the Copyright Office for denying his application to register a visual artwork that he claims was authored "autonomously" by an AI program. Dr. Thaler argued that human authorship is not required by the Copyright Act. In August 2023, a U.S. district court granted summary judgment in favor of the Copyright Office. The court held that "human authorship is an essential part of a valid copyright claim," reasoning that only human authors need copyright as an incentive to create expressive works.

In March 2025, the U.S. Court of Appeals for the D.C. Circuit affirmed the district court's decision in *Thaler v. Perlmutter*, holding that the Copyright Act "requires all eligible work to be authored in the first instance by a human being." The court reasoned that several provisions of the Copyright Act imply that it uses the word "author" only to refer to human beings, including provisions (1) vesting copyright ownership "initially in the author"; (2) limiting copyright duration to 70 years after "the author's death"; (3) providing for inheritance of certain rights by the author's "widow or widower" or "surviving children or grandchildren"; (4) requiring a signature to transfer copyright ownership; (5) extending protection to unpublished works regardless of the author's "nationality or domicile"; and (6) defining a "joint work" based on the authors' "intention" to merge their contributions in a certain way.

In addition, the D.C. Circuit observed that the Copyright Office had adopted the human-authorship requirement before Congress enacted the current Copyright Act. The court thus inferred that Congress meant to adopt the human-authorship requirement when it enacted the law. Based on its holding that the Copyright Act requires human authorship, the court found it unnecessary to consider the Copyright Office's argument that the Constitution's IP Clause also requires human authorship. On May 12, 2025, the court denied Dr. Thaler's petition to rehear the case en banc (i.e., by all of the court's judges).

May Humans Copyright Works That They Create Using AI?

Assuming that copyrightable works require a human author, works created by humans with the assistance of generative AI might be entitled to copyright protection depending on the nature of human involvement in the creative process. As discussed below, the Copyright Office has sought to delineate what authors must do to satisfy the human-authorship requirement when using generative AI.

In March 2023, the Copyright Office released Copyright Registration Guidance regarding "works containing material generated by [AI]" (the AI Guidance). Granting that human authors may use AI in the creative process, the AI Guidance states that "what matters is the extent to which the human had creative control over the work's expression." Thus, the AI Guidance states, when AI "determines the expressive elements of its output, the generated material is not the product of human authorship." On the other hand, works containing AI-generated material may be copyrighted under some circumstances, such as "sufficiently creative" human arrangements or modifications of AI-generated material or works that combine AI-generated and human-authored material. The AI Guidance states that authors may claim copyright protection only "for their own contributions" to such works, and they must identify and disclaim AI-generated parts of the works when applying to register their copyright.

Three copyright registration denials highlighted by the Copyright Office illustrate that, in general, the office will not find human authorship where an AI program generates works in response to user prompts:

- *Zarya of the Dawn*: A February 2023 decision that AI-generated illustrations for a graphic novel were not copyrightable, although the human-authored text of the novel and overall selection and arrangement of the images and text in the novel could be copyrighted.

- *Théâtre D'opéra Spatial*: A September 2023 decision that an artwork generated by AI and then modified by the applicant could not be copyrighted, since the applicant failed to identify and disclaim the AI-generated portions as required by the AI Guidance.

- *SURYAST*: A December 2023 decision that an artwork generated by an AI system combining a "base image" (an original photo taken by the applicant) and a "style image" the applicant selected (Vincent van Gogh's *The Starry Night*) could not be copyrighted, since the AI system was "responsible for determining how to interpolate [i.e., combine] the base and style images."

While the Copyright Office's decisions indicate that it may not be possible to obtain copyright protection for many AI-generated works, the issue remains unsettled. An applicant may file suit in U.S. district court to challenge the Copyright Office's final decision to refuse to register a copyright. The putative author of *Théâtre D'opéra Spatial*, for instance, has sued the Copyright Office for declining to register that work. While the Copyright Office notes that courts sometimes give weight to the office's experience and expertise, courts are not legally bound to adopt the office's interpretations of the Copyright Act, such as its application of the authorship requirement to AI-assisted works.

In January 2025, the Copyright Office published the part of its *Copyright and Artificial Intelligence* report addressing the copyrightability of AI-generated works. Reinforcing the AI Guidance's emphasis on "creative control," the report concludes that, "given current generally available technology, prompts alone do not provide sufficient human control to make users of an AI system the authors of the output." The report contends that the Copyright Act's distinction between copyrightable "works" and noncopyrightable "ideas" precludes copyrightability for works generated by AI in response to user prompts. Specifically, the report argues, "[p]rompts essentially function as instructions that convey unprotectible ideas" and "do not control how the AI system processes them in generating the output."

Some commentators assert that certain AI-generated works should receive copyright protection, comparing AI programs to other tools human authors have used to create copyrighted works. For example, the U.S. Supreme Court held in the 1884 case *Burrow-Giles Lithographic Co. v. Sarony* that photographs can be entitled to copyright protection where the photographer makes decisions regarding creative elements such as composition, arrangement, and lighting. Some copyright applicants argue that generative AI programs may function as tools, analogous to cameras. The Copyright Office disputes the photography analogy, arguing that AI users do not exercise sufficient control to characterize generative AI as a tool used by an author. Instead, the Copyright Office has analogized an AI user to "a client who hires an artist" and gives that artist only "general directions." At least one applicant's attorney has argued that the Copyright Act does not require such exacting creative control, observing that certain photographs and modern art incorporate a degree of happenstance.

Regarding works that combine human-authored and AI-generated material, the Copyright Office reports that, in the time since it issued the AI Guidance, it "has registered hundreds of works that incorporate AI-generated material, with the registration covering the human author's contribution to the work." The office contends that new legislation regarding "the copyrightability of AI-generated material" is currently not needed, indicating that courts "will provide further guidance on the human authorship requirement as it applies to specific uses of AI" and that, since each work must be analyzed individually, "greater clarity would be difficult to achieve" through legislation.

Copyright Infringement by Generative AI Programs

Does the AI Training Process Infringe Copyrights in Other Works?

AI systems are trained to create literary, visual, and other works by exposure to large amounts of data, which may include text, images, and other works downloaded from the internet or otherwise obtained by AI companies. This process often involves making digital copies of existing works. Copyright owners have filed several dozen lawsuits claiming that creating these digital copies without permission to train AI systems infringes their exclusive right to make reproductions (or copies).

Many AI companies and some legal scholars argue that using copyrighted works to train AI systems constitutes fair use and is therefore noninfringing. Whether or not unauthorized copying constitutes fair use depends on four nonexclusive factors that Congress set forth in the Copyright Act:

1. the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes;
2. the nature of the copyrighted work;
3. the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and
4. the effect of the use upon the potential market for or value of the copyrighted work.

The Supreme Court has stated that fair use is a "flexible" doctrine, and "its application may well vary depending upon context." As to the first factor, the Court has held that some uses with a "transformative" purpose (such as parodies) may be fair, although in 2023 it cautioned that transformativeness is "a matter of degree." The Court has described the fourth factor as the "most important" one.

In June 2025, the U.S. District Court for the Northern District of California issued summary judgment decisions in two lawsuits concerning generative AI and fair use. In one case, *Bartz v. Anthropic PBC*, the defendant had trained an AI system (Claude) on books obtained through a combination of buying print copies and downloading digital copies from "pirate sites." The district court held that it was fair use to copy the books to train Claude, reasoning that generative AI is "quintessentially transformative" (factor 1), that copying entire books was "reasonably necessary" to train Claude (factor 3), and that Claude does not create copyright-infringing outputs that would "displace demand" for the books (factor 4). The court found that only factor 2 tilted against fair use, since the books were used for their "expressive qualities." The court also held that it was fair use for Anthropic to convert the print books it bought into digital versions. But, the court held, it was not fair use for Anthropic to download pirated books and maintain them in a "central library," only parts of which were used for AI training. On July 17, the court certified a class action for copyright owners of books Anthropic had downloaded from certain pirate libraries.

In the other case, *Kadrey v. Meta Platforms, Inc.*, a different judge ruled that, based on the record before the court, it was fair use for Meta to train a generative AI system (Llama) on books it downloaded from pirated "shadow libraries." As in *Bartz*, the court in *Kadrey* held that this use was "highly transformative" and that factor 1 as well as factor 3 favored fair use, while factor 2 did not. Regarding factor 4, the court criticized plaintiffs for failing to develop evidence of market harm, potentially forfeiting a winning argument. The court opined in dicta that factor 4 might weigh decisively against fair use in other generative AI cases. Disagreeing with the judge in *Bartz*, the court in *Kadrey* reasoned that market harms from generative AI are not limited to copyright-infringing outputs but may also include noninfringing outputs that compete with copyrighted works (for example, because they concern the same topic).

The *Bartz* and *Kadrey* judges' disagreement reflects a wider debate over whether "market dilution" from noninfringing AI outputs is a cognizable harm under factor 4. The Copyright Office, for instance, has argued that market dilution may weigh against fair use, while some scholars have criticized this view.

The *Bartz* and *Kadrey* decisions also took different approaches to the initial downloading of plaintiffs' books without the authors' permission. Unlike in *Bartz*, the court in *Kadrey* did not conduct a separate fair-use analysis of these downloads. The court reasoned that Meta's purpose in downloading the books was to train Llama, and "[b]ecause Meta's ultimate use of the plaintiffs' books was transformative, so too was Meta's downloading . . . ." By contrast, the court in *Bartz* found that Anthropic had impermissibly obtained "a central library of works to be available for any number of further uses." Yet in dicta, the *Bartz* court doubted that it would ever be fair use to download books from pirate sites instead of purchasing them. On July 14, Anthropic filed a motion seeking an appeal or reconsideration of the court's decision in *Bartz*, arguing that the decisions in *Bartz* and *Kadrey* "cannot be reconciled" on the downloading issue.

Given the fact-specific nature of fair use, courts in other generative AI cases may conduct their own analyses and reach different answers as to fair use and its constituent factors. In May 2025, the Copyright Office released a prepublication version of the Generative AI Training part of its *Copyright and Artificial Intelligence* report, which concluded that "it is not possible to prejudge litigation outcomes" and that "some uses of copyrighted works for generative AI training will qualify as fair use, and some will not."

Do AI Outputs Infringe Copyrights in Other Works?

Some outputs of AI programs might infringe copyrights in other works they resemble that were used to train the AI. Copyright owners may be able to establish that such outputs infringe their copyrights if the AI program both (1) had access to their works and (2) created "substantially similar" outputs. First, to establish copyright infringement, a plaintiff must prove the infringer "actually copied" the underlying work. This element is sometimes proven circumstantially by evidence that the infringer "had access to the work." For AI outputs, access might be shown by evidence that the AI program was trained using the underlying work. Such evidence might show, for instance, that a copy of the underlying work was located on an internet site that was downloaded or "scraped" to train the AI program.

Second, a plaintiff must prove that the new work is "substantially similar" to the copyrighted work to establish infringement. The substantial similarity test is difficult to define. Courts have variously described the test as requiring, for example, that the works have "a substantially similar total concept and feel" or "overall look and feel" or that "the ordinary reasonable person would fail to differentiate between the two works." Leading cases have also stated that this determination considers both "the qualitative and quantitative significance of the copied portion in relation to the plaintiff's work as a whole." For AI-generated outputs, no less than for traditional works, the "substantial similarity" analysis may require courts to make these kinds of comparisons between the output and the plaintiff's work.

OpenAI has argued that "[w]ell-constructed AI systems generally do not regenerate, in any nontrivial portion, unaltered data from any particular work in their training corpus." Thus, according to OpenAI, copyright-infringing outputs would be "an unlikely accidental outcome" of such systems. One study found "a significant amount of copying" in less than 2% of the images created by Stable Diffusion, though the authors claimed that their methodology "likely underestimates the true rate" of copying.

Two kinds of AI outputs may raise special concerns. First, some AI programs may be used to create works involving existing fictional characters. These works may run a heightened risk of infringement, since characters sometimes enjoy copyright protection distinct from the specific works in which they appear. Second, some AI programs may be prompted to create works "in the style of" a particular artist or author, although some AI programs may now be designed to "decline" such prompts. These outputs are not necessarily infringing, as copyright law generally protects only against the copying of specific works. For example, a song generated by AI in the style and simulated voice of a human performer might not infringe any copyright, although voice simulations may potentially violate some state right-of-publicity laws. As a separate issue, judges and commentators disagree about whether market dilution from stylistically similar outputs could weigh against a fair-use defense for training AI on an author's work, as discussed above.

If a generative AI output infringes a copyright in an existing work, both the AI user and the AI company could potentially be liable under current law. For instance, the user might be directly liable for prompting the AI program to generate an infringing output. It may be challenging to analyze the user's liability in some cases, since the user might not have direct access to—or even be aware of—a copyrighted work purportedly infringed by an AI output. The AI company could also potentially face liability under the doctrine of "vicarious infringement." Vicarious infringement applies to defendants who have "the right and ability to supervise the infringing activity" and "a direct financial interest in such activities."

Considerations for Congress

Congress may consider whether to address any of the copyright law questions raised by generative AI programs through amendments to the Copyright Act or other legislation. Congress may, for example, consider legislation clarifying whether AI-generated works are copyrightable or under what circumstances the process of training generative AI programs may constitute fair use. Alternatively, given the limited time courts have had to address these issues, Congress may adopt a wait-and-see approach. As courts decide more cases involving generative AI, they may be able to provide greater guidance and predictability in this area. Based on such outcomes, Congress may reassess whether legislation is needed.

Congress may also consider the practical implications of requiring AI companies to identify, seek permission from, or compensate copyright owners should court decisions or future legislation determine that training generative AI systems is not a fair use of copyrighted works. Commentators have debated whether it is feasible to require companies to identify and pay owners of the large number of works needed to train AI systems, as well as whether the value of such compensation to owners would be outweighed by transaction or administration costs. One scholar, acknowledging that "[i]t would . . . be impossible for an AI developer to identify and clear billions of rights claims on an individual basis," argues that it may be feasible instead to create markets for AI training data via means such as content aggregation (e.g., TV streaming services), collective management organizations (or CMOs, such as those that manage rights to musical works), compulsory licenses (which exist for certain uses of sound recordings), or technological measures (e.g., giving rightsholders a means to opt out of using works for AI training). Another scholar questions the practicality of CMOs for AI training data, in part due to the large volume and variety of works used to train AI, while observing that some "large rights holders" (such as the Associated Press) have individually contracted with AI companies for the use of their works.

In the Generative AI Training part of its *Copyright and Artificial Intelligence* report, the Copyright Office contends that voluntary licensing—either by individual rightsholders or through CMOs—is sometimes (though perhaps not always) feasible for licensing copyrighted works to train AI. For instance, the report indicates that "[s]ome AI systems have now been trained exclusively on licensed or public domain works." The report expresses normative and practical reservations about nonvoluntary approaches to licensing training data, such as compulsory licenses or requiring copyright holders to "opt out" if they do not consent to the use of their works to train AI. Thus, the report "recommends allowing the licensing market to continue to develop without government intervention" for now. Congress may consider these recommendations as well as differing perspectives such as those surveyed in the report.

Author Information

Christopher T. Zirpoli
Legislative Attorney

Disclaimer

This document was prepared by the Congressional Research Service (CRS). CRS serves as nonpartisan shared staff to congressional committees and Members of Congress. It operates solely at the behest of and under the direction of Congress. Information in a CRS Report should not be relied upon for purposes other than public understanding of information that has been provided by CRS to Members of Congress in connection with CRS's institutional role. CRS Reports, as a work of the United States Government, are not subject to copyright protection in the United States. Any CRS Report may be reproduced and distributed in its entirety without permission from CRS. However, as a CRS Report may include copyrighted images or material from a third party, you may need to obtain the permission of the copyright holder if you wish to copy or otherwise use copyrighted material.