Congressional Research Service
Informing the legislative debate since 1914

Legal Sidebar

Generative Artificial Intelligence and Copyright Law

February 24, 2023

Recent innovations in artificial intelligence (AI) are raising new questions about how copyright law principles such as authorship, infringement, and fair use will apply to content created or used by AI. So-called "generative AI" computer programs—such as Open AI's DALL-E 2 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 prompts (or "inputs"). These generative AI programs are "trained" to generate such works partly by exposing them to large quantities of existing works such as writings, photos, paintings, and other artworks. This Legal Sidebar explores questions that courts and the U.S. Copyright Office have begun to confront regarding whether the outputs of generative AI programs are entitled to copyright protection as well as how training and using these programs might infringe copyrights in other works.

Copyright in Works Created with Generative AI

The widespread use of generative AI programs raises the question of who, if anyone, may hold the copyright to works created using these programs, given that the AI's user, the AI's programmer, and the AI program itself all play a role in the creation of these works.

Do AI Outputs Enjoy Copyright Protection?

The question of whether or not copyright protection may be afforded to AI outputs—such as images created by DALL-E or texts created by ChatGPT—is likely to hinge partly on the concept of "authorship." The Copyright Act generally affords copyright protection to "original works of authorship." Although the Copyright Act does not define who (or what) may be an "author," the U.S Copyright Office recognizes copyright only in works "created by a human being." Courts have likewise refused to afford copyright protection to non-human authors—for example, a monkey who took a series of photos.

A recent lawsuit has challenged the human-authorship requirement in the context of works purportedly "authored" by AI. In June 2022, Stephen Thaler sued the Copyright Office for denying an application to register a visual artwork that he claims was authored by an AI program called the Creativity Machine. Dr.

Congressional Research Service
https://crsreports.congress.gov
LSB10922

Thaler asserts the picture was created "autonomously by machine," and he argues that human authorship is not required by the Copyright Act. The lawsuit is pending.

Even assuming that a copyrightable work requires a human author, works created using generative AI could arguably be entitled to copyright protection, depending on the nature of human involvement in the creative process. A recent copyright proceeding shows that the Copyright Office may be skeptical of this argument. In September 2022, writer Kristina Kashtanova registered a copyright for her graphic novel, which was illustrated entirely with images generated by Midjourney in response to her text inputs. In October, however, the Copyright Office initiated cancellation proceedings, noting that Ms. Kashtanova had not disclosed her use of AI. Ms. Kashtanova responded by arguing that she authored the images via "a creative, iterative process" entailing "multiple rounds of composition, selection, arrangement, cropping, and editing for each image." She contrasted her creative process with the "autonomously generated" image Dr. Thaler tried to register. Nevertheless, on February 21, 2023, the Copyright Office determined that the images did not have a human author and therefore were not copyrightable.

Some commentators assert that at least some AI-generated works should receive copyright protection, arguing that AI programs are analogous to other tools that human beings have used to create copyrighted works. For example, the Supreme Court has held since 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. Generative AI programs might be seen as another tool, akin to camera equipment, that can be used by human authors. Ms. Kashtanova argued that her own creative process was similar to that of a photographer.

Other commentators and the Copyright Office, disputing the photography analogy, question whether AI users exercise sufficient creative control for AI to be considered merely a tool. In Ms. Kashtanova's case, the Copyright Office reasoned that, instead of being "a tool that Ms. Kashtanova controlled and guided to reach her desired image, Midjourney generates images in an unpredictable way." The Copyright Office thus compared the AI user to "a client who hires an artist" as opposed to an artist in her own right. Similarly, one law professor has suggested that the AI user creates only a non-copyrightable, abstract "idea" rather than a copyrightable work expressing the idea: "If I ask Dall-E to produce a painting of hedgehogs having a tea party on the beach, I have contributed nothing more than an idea." According to this argument, the artwork lacks an author and is thus in the public domain.

In sum, it is difficult to predict whether copyright law will recognize any protection for generative AI outputs. While the Copyright Office's handling of Ms. Kashtanova's case to date indicates skepticism toward copyright protection for AI-generated works, the Copyright Office does not have the only or the final say on U.S. copyright law. For example, applicants may request reconsideration of adverse decisions by the Copyright Office and may ultimately challenge those decisions in U.S. district court, as Dr. Thaler has done. It remains to be seen how federal courts will handle such challenges and other cases concerning copyright in AI-generated works.

Who Owns the Copyright to Generative AI Outputs?

Assuming some AI-created works may be eligible for copyright protection, who owns that copyright? In general, the Copyright Act vests ownership "initially in the author or authors of the work." Given the lack of judicial or Copyright Office decisions recognizing copyright in AI-created works to date, however, no clear rule has emerged identifying who the "author or authors" of these works could be. Returning to the photography analogy, the AI's creator might be compared to the camera maker, while the AI user who prompts the creation of a specific work might be compared to the photographer who uses that camera to capture a specific image. On this view, the AI's user would be considered the author and, therefore, the initial copyright owner. The creative choices involved in coding and training the AI, on the other hand, might give an AI's creator a stronger claim to some form of authorship than the manufacturer of a camera.

Regardless of who may be the initial copyright owner of an AI output, companies that provide AI software may attempt to allocate the respective ownership rights of the company and its users via contract, such as the company's terms of service. OpenAI's current Terms of Use, for example, appear to assign any copyright to the user: "OpenAI hereby assigns to you all its right, title and interest in and to Output." A previous version of these terms, by contrast, purported to give OpenAI such rights. Either way, OpenAI does not seem to address who would own the copyright in the absence of such terms. As one scholar commented, OpenAI appears to "bypass most copyright questions through contract."

Copyright Infringement by Generative AI

Generative AI also raises questions about copyright infringement. Commentators and courts have begun to address whether generative AI programs may infringe copyright in existing works, either by making copies of existing works to train the AI or by generating outputs that resemble those existing works.

Does the AI Training Process Infringe Copyright in Other Works?

AI systems are "trained" to create literary, visual, and other artistic works by exposing the program to large amounts of data, which may consist of existing works such as text and images from the internet. This training process may involve making digital copies of existing works, carrying a risk of copyright infringement. As the U.S. Patent and Trademark Office has described, this process "will almost by definition involve the reproduction of entire works or substantial portions thereof." OpenAI, for example, acknowledges that its programs are trained on "large, publicly available datasets that include copyrighted works" and that this process "necessarily involves first making copies of the data to be analyzed." Creating such copies, without express or implied permission from the various copyright owners, may infringe the copyright holders' exclusive right to make reproductions of their work.

AI companies may argue that their training processes constitute fair use and are therefore non-infringing. Whether or not copying constitutes fair use depends on four statutory factors under 17 U.S.C. § 107:

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.

Some stakeholders argue that the use of copyrighted works to train AI programs should be considered a fair use under these factors. Regarding the first factor, OpenAI argues its purpose is "transformative" as opposed to "expressive" because the training process creates "a useful generative AI system." OpenAI also contends that the third factor supports fair use because the copies are not made available to the public but are used only to train the program. For support, OpenAI cites *The Authors Guild, Inc. v. Google, Inc.*, in which the U.S. Court of Appeals for the Second Circuit held that Google's copying of entire books to create a searchable database that displayed excerpts of those books constituted fair use.

These arguments may soon be tested in court, as plaintiffs have recently filed multiple lawsuits alleging copyright infringement via AI training processes. On January 13, 2023, several artists filed a putative class action lawsuit alleging their copyrights were infringed in the training of AI image programs, including Stable Diffusion. The class action lawsuit claims that defendants "downloaded or otherwise acquired copies of billions of copyrighted images without permission" to use as "training images," making and storing copies of those images without the artists' consent. Similarly, on February 3, 2023, Getty Images filed a lawsuit alleging that "Stability AI has copied at least 12 million copyrighted images

from Getty Images' websites ... in order to train its Stable Diffusion model." Both lawsuits appear to dispute any characterization of fair use, arguing that Stable Diffusion is a commercial product, weighing against fair use under the first statutory factor, and that the program undermines the market for the original works, weighing against fair use under the fourth factor.

Do AI Outputs Infringe Copyrights in Other Works?

AI programs might also infringe copyright by generating outputs that resemble existing works. Under U.S. case law, copyright owners may be able to show 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 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. For instance, the underlying work might be part of a publicly accessible internet site that was downloaded or "scraped" to train the AI program.

Second, a plaintiff must prove the new work is "substantially similar" to the underlying work to establish infringement. The substantial similarity test is difficult to define and varies across U.S. courts. Courts have variously described the test as requiring, e.g., 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 traditional works, the "substantial similarity" analysis may require courts to make these kinds of comparisons between the AI output and the underlying work.

There is significant disagreement as to how likely it is that generative AI programs will copy existing works in their outputs. OpenAI argues 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, OpenAI states, infringement "is an unlikely accidental outcome." By contrast, the Getty Images lawsuit alleges that "Stable Diffusion at times produces images that are highly similar to and derivative of the Getty Images." One study has found "a significant amount of copying" in a small percentage (less than 2%) of the images created by Stable Diffusion. Yet the other, class action lawsuit against Stable Diffusion appears to argue that *all* Stable Diffusion outputs are potentially infringing, alleging that they are "generated exclusively from a combination of ... copies of copyrighted images."

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 copyright infringement insofar as characters sometimes enjoy copyright protection in and of themselves. Second, some AI programs may be used to create artistic or literary works "in the style of" a particular artist or author. These outputs are not necessarily infringing, as copyright law generally prohibits the copying of specific works rather than an artist's overall style. Nevertheless, some artists are concerned that generative AI programs are uniquely capable of mass-producing works that copy their style, potentially undercutting the value of their work. In the class action lawsuit against Stable Diffusion, for example, plaintiffs claim that few human artists can successfully mimic another artist's style, whereas "AI Image Products do so with ease."

A final question is who is (or should be) liable if generative AI outputs do infringe copyrights in existing works. Under current doctrines, both the AI user and the AI company could potentially be liable. For instance, even if a user were directly liable for infringement, the AI company could potentially face liability under the doctrine of "vicarious infringement," which applies to defendants who have "the right and ability to supervise the infringing activity" and "a direct financial interest in such activities." The class action lawsuit against Stable Diffusion, for instance, claims that the defendant AI companies are

vicariously liable for copyright infringement. One complication of AI programs is that the user might not be aware of—or have access to—a work that was copied in response to the user's prompt. Under current law, this may make it challenging to analyze whether the user is liable for copyright infringement.

Considerations for Congress

Congress may wish to consider whether any of the copyright law questions raised by generative AI programs require amendments to the Copyright Act or other legislation. Congress may, for example, wish to consider legislation clarifying whether AI-generated works are copyrightable, who should be considered the author of such works, or when the process of training generative AI programs constitutes fair use.

Given how little opportunity the courts and Copyright Office have had to address these issues, Congress may wish to adopt a wait-and-see approach. As the courts gain experience handling cases involving generative AI, they may be able to provide greater guidance and predictability in this area through judicial opinions. Based on the outcomes of early cases in this field, such as those summarized above, Congress may reassess whether legislative action is needed.

Author Information

Christopher T. Zirpoli
Legislative Attorney

Disclaimer