

Generative Artificial Intelligence and Copyright Law

Updated May 11, 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 outputs 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 content 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—likely hinges at least partly on the concept of “authorship.” The U.S. Constitution [authorizes](#) 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.” Although the Constitution and Copyright Act do not explicitly 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 declined to extend copyright protection to nonhuman authors. For example, appellate courts have held in various cases that [a monkey](#) who took a series of photos lacked standing to sue under the Copyright Act; that some human creativity was required to copyright a book purportedly inspired by [celestial beings](#); and that a [living garden](#) could not be copyrighted as it lacked a human author.

Congressional Research Service

<https://crsreports.congress.gov>

LSB10922

Pocket Constitution



The Declaration of Independence
The Constitution of the United States
The Bill of Rights
Amendments XI–XXVII
Gettysburg Address



TheCapitolNet

TCNFPC.com

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. 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.

Assuming that a copyrightable work requires a human author, works created by humans using generative AI could arguably be entitled to copyright protection, depending on the nature of human involvement in the creative process. However, a recent [copyright proceeding](#) and subsequent [Copyright Registration Guidance](#) indicate that the Copyright Office is unlikely to find the requisite human authorship where an AI program generates works in response to simple text prompts. In September 2022, Kris Kashtanova registered a copyright for a graphic novel they illustrated with images generated by Midjourney in response to textual inputs. In October, the Copyright Office initiated cancellation proceedings, [noting](#) that Kashtanova had not disclosed their use of AI. Kashtanova [responded](#) by arguing that they authored the images via “a creative, iterative process,” contrasting this process with the image Dr. Thaler tried to register. Nevertheless, on February 21, 2023, the Copyright Office [determined](#) that the images were not copyrightable, deciding that Midjourney, rather than Kashtanova, authored the “visual material.” Building on this decision, the Copyright Office released guidance in March [stating](#) that, when AI “determines the expressive elements of its output, the generated material is not the product of human authorship” (and therefore 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 a camera, that can be used by human authors to create copyrightable works, as Kashtanova [argued](#).

Other [commentators](#) and the Copyright Office dispute the photography analogy and question whether AI users exercise sufficient creative control for AI to be considered merely a tool. In Kashtanova’s case, the Copyright Office [reasoned](#) that, rather than “a tool that [] Kashtanova controlled and guided to reach [their] desired image, Midjourney generates images in an unpredictable way.” The Copyright Office instead [compared](#) the AI user to “a client who hires an artist” to create something and provides only “general directions.” The office’s March 2023 guidance similarly [claims](#) that “users do not exercise ultimate creative control over how [current generative AI] systems interpret prompts and generate materials.” One of Kashtanova’s lawyers, on the other hand, [argues](#) that the Copyright Act does not require such exacting creative control, noting that certain kinds of photography and visual art incorporate some degree of happenstance.

Some commentators argue that the Copyright Act’s distinction between copyrightable “works” and noncopyrightable “[ideas](#)” supplies another reason that copyright should not protect AI-generated works. One law professor has [suggested](#) that the human user who enters a text prompt into an AI program—for instance, asking DALL-E “to produce a painting of hedgehogs having a tea party on the beach”—has “contributed nothing more than an idea” to the finished work. According to this [argument](#), the output image lacks a human author and cannot be copyrighted.

While the Copyright Office’s actions to date indicate that it may be challenging to obtain copyright protection for AI-generated works, the issue remains unsettled. Applicants may file suit in U.S. district court to [challenge](#) the Copyright Office’s final decisions to refuse to register a copyright (as Dr. Thaler has done), and it remains to be seen what federal courts will decide concerning whether AI-generated works may be copyrighted. While the Copyright Office [notes](#) that courts sometimes give weight to the office’s experience and expertise in this field, courts will not necessarily adopt the office’s interpretations

of the Copyright Act. In addition, the Copyright Office’s guidance accepts that works “[containing](#)” AI-generated material may be copyrighted under some circumstances, such as “sufficiently creative” human arrangements or modifications of that material.

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 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 noninfringing. 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.

Regarding the fourth fair use factor, some generative AI applications have raised concern that training AI programs on copyrighted works allows them to generate works that compete with the original works. For example, an AI-generated song called “[Heart on My Sleeve](#),” made to sound like the artists Drake and The Weeknd, was heard millions of times in April 2023 before it was removed by various streaming services. Universal Music Group, which has deals with both artists, [argues](#) that AI companies violate copyright by using these artists’ songs in training data.

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 Midjourney and 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, 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 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. Regarding the AI-generated song “Heart on My Sleeve,” for instance, one commentator notes that the imitation of Drake or another artist’s voice appears not to violate copyright law provided that the song does not copy an “individual existing work” (e.g., the lyrics or melodies of a particular Drake song), although it may raise concerns under some states’ right-of-publicity laws. 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

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.

Learn how Capitol Hill really works

All of our programs and any combination of their topics can be tailored for custom training for your organization.

For more than 40 years, TheCapitol.Net and its predecessor, Congressional Quarterly Executive Conferences, have been teaching professionals from government, military, business, and NGOs about the dynamics and operations of the legislative and executive branches and how to work with them.

Our training, on-site and online, and publications include congressional operations, legislative and budget process, communication and advocacy, media and public relations, research, testifying before Congress, legislative drafting, critical thinking and writing, and more.

- **Diverse Client Base**—We have tailored hundreds of custom on-site and online training programs for Congress, numerous agencies in all federal departments, the military, law firms, lobbying firms, unions, think tanks and NGOs, foreign delegations, associations and corporations, delivering exceptional insight into how Washington works.™
- **Experienced Program Design and Delivery**—We have designed and delivered hundreds of custom programs covering congressional/legislative operations, budget process, media training, writing skills, legislative drafting, advocacy, research, testifying before Congress, grassroots, and more.
- **Professional Materials**—We provide training materials and publications that show how Washington works. Our publications are designed both as course materials and as invaluable reference tools.
- **Large Team of Experienced Faculty**—More than 150 faculty members provide independent subject matter expertise. Each program is designed using the best faculty member for each session.
- **Non-Partisan**—TheCapitol.Net is non-partisan.
- **GSA Schedule**—TheCapitol.Net is on the GSA Schedule for custom training: GSA Contract GS02F0192X.

Please see our Capability Statement on our web site at TCNCS.com.

Custom training programs are designed to meet your educational and training goals, each led by independent subject-matter experts best qualified to help you reach your educational objectives and align with your audience.

As part of your custom program, we can also provide online venue, classroom space, breaks and meals, receptions, tours, and online registration and individual attendee billing services.

For more information about custom on-site training for your organization, please see our web site: TCNCustom.com or call us: 202-678-1600, ext 115.



TheCapitol.Net

Non-partisan training and publications that show how Washington works.™

PO Box 25706, Alexandria, VA 22313-5706
202-678-1600 • www.thecapitol.net



TheCapitol.Net is
on the GSA Schedule
for custom training.
GSA Contract GS02F0192X

