

Who Is Responsible for AI Copyright Infringement?

Twenty-one-year-old college student Shane hopes to write a song for his boyfriend. In the past, Shane would have had to wait for inspiration to strike, but now he can use generative artificial intelligence to get a head start. Shane decides to use Anthropic’s AI chat system, Claude, to write the lyrics. Claude dutifully complies and creates the words to a love song. Shane, happy with the result, adds notes, rhythm, tempo, and dynamics. He sings the song and his boyfriend loves it. Shane even decides to post a recording to YouTube, where it garners 100,000 views.

But Shane did not realize that this song’s lyrics are similar to those of “Love Story,” Taylor Swift’s hit 2008 song. Shane must now contend with copyright law, which protects original creative expression such as music. Copyright grants the rights owner the exclusive rights to reproduce, perform, and create derivatives of the copyrighted work, among other things. If others take such actions without permission, they can be liable for damages up to \$150,000. So Shane could be on the hook for tens of thousands of dollars for copying Swift’s song.

Copyright law has surged into the news in the past few years as one of the most important legal challenges for generative AI tools like Claude—not for the output of these tools but for how they are trained. Over two dozen pending court cases grapple with the question of whether training generative AI systems on copyrighted works without compensating or getting permission from the creators is lawful or not. Answers to this question will shape a burgeoning AI industry that is predicted to be worth \$1.3 trillion by 2032.

Yet there is another important question that few have asked: *Who* should be liable when a generative AI system creates a copyright-infringing output? Should the user be on the hook? Shane only requested a generic love song, not “Love Story” or one like it. What about the provider of the AI tool? It is not in Anthropic’s interest for Claude to produce an infringing song, and the company likely took measures to avoid infringing outputs.

I propose that neither option is desirable. Instead, the AI system itself—as a fictitious legal person—should be the copyright infringer. Any human liability for infringement should be determined on a more nuanced secondary liability basis. While this is a viable approach under copyright doctrine, it also suggests that we can more broadly reimagine the role of AI as the perpetrator to respond to the unique legal conundrums posed by increasingly autonomous systems.

Who caused the machine to infringe?

Some AI-generated outputs will undoubtedly infringe copyright. Scholars including Matthew Sag, James Grimmelman, and Timothy Lee have already shown how generative AI systems can produce images that infringe on copyrighted content such as *Peanuts*’ cartoon beagle Snoopy, Nintendo’s Italian plumber hero Mario, and Banksy’s iconic *Girl with Balloon* mural. In March, the Guangzhou Internet Court in China issued the first ruling involving copyright infringement and generative AI, finding that the output at issue infringed the copyright for Japanese superhero Ultraman. (In that case, the court found the AI provider liable for the infringing output.)

The question of who should be liable is especially challenging because AI systems are black boxes. Developers and users are not able to precisely predict or explain why a particular output occurs. The US Copyright Office has rejected copyright registrations for AI-generated works on the basis that, due to the black box, the AI system—not the developer or user—is creating the specific expressive output. The US Patent and Trademark Office has similarly rejected the notion that an AI system can be an inventor for purposes of patent law. The unpredictability and semiautonomous nature of AI has led many commentators to suggest that a paradigm shift in the law is needed for the AI era.

This is not the first time copyright law has encountered difficult questions of liability in the face of emerging technologies. From the printing press to search engines, complex machines have posed new questions for who should be liable for resulting infringements. In most cases, if infringement occurred, the infringer has usually been obvious. Whether it was the painter of the art, author of the book, or copier of the music, there's been a clear line from infringer to infringement. In these *corporeal infringement* cases, the identity of

Users or developers?

I propose that infringement carried out through generative AI—which I term *artificial infringement*—is merely the third stage in this liability evolution. Courts can use the volition requirement, which helped them solve the liability challenges of mechanical infringement, to determine who should be held liable for AI-generated infringements as well.

There are scenarios where it seems fairly likely the user or the developer caused the AI system to infringe. Users may have specifically tried to make the AI system generate infringing outputs or have even engaged in adversarial machine learning aimed at circumventing the system's safeguards against infringement. It is also possible that the developer of the AI system offered a system that is almost guaranteed to infringe most of the time, either due to malintent or poor design.

But in most artificial infringement cases, it will be difficult to show that a human acted with volition. The creation and launch of an AI system involves many individuals. Most providers of AI systems actively seek to limit instances of infringing outputs. This is why, for

The AI system itself—as a fictitious legal person—should be the copyright infringer. Any human liability for infringement should be determined on a more nuanced secondary liability basis.

the infringer was a foregone conclusion because the infringement occurred by the person's own hand.

As machines became more complex, however, multiple parties became involved in the copying of copyrighted works, which defied this straightforward understanding of liability. For example, posting infringing content online involved the user who posted the content, their internet service provider, the website to which it was posted, and the internet service provider for that website. A remote digital video recording device for taping television programs required both the system that facilitated the recording and the user who selected the specific programming for later viewing. In these and other *mechanical infringement* cases, courts introduced a new term to copyright law: *volition*. Volition asks who willed or caused the infringement to occur—or, as one court put it, “who actually presses the button”? Although volition was always part of copyright law in the background, courts had to bring it to the fore to address the complications of mechanical infringement. With the aid of volition, courts in both these examples determined that the user was the infringer because the user—not the service provider—caused the infringement to occur.

example, you cannot prompt some generative AI systems to create an output in the style of well-known artists such as Pablo Picasso or Salvador Dalí. Users also generally do not intend infringing works to be generated or, like Shane, are unaware when outputs are infringing.

We could ignore this complex reality and make the AI system provider or user automatically liable for all infringements. But this is undesirable for several reasons. First, such a rule would ignore legal precedent that, without some causation or intent, providers are generally not liable merely for offering a product or automatic service that has substantial noninfringing uses. Second, the Copyright Office has already refused to register copyrights in AI-generated outputs because the developer and user do not have sufficient control over the expression, which also forms the basis for the infringement. Third, imposing liability on developers for all resulting infringements despite their best intentions could deter new market entrants, inhibiting robust competition and solidifying the positions of wealthy market leaders such as OpenAI and Meta. Finally, absolute liability would not necessarily deter infringement because of the lack of precise control over the AI black box.

AI systems as copyright infringers

A proper volition analysis instead suggests a novel solution: courts should consider the AI system, as an artificial legal person, to be the infringer. The AI system is what is determining the expression in a particular output. While the developer provides the infrastructure and the user the prompt, the actual determination of whether an output will contain infringing content is made inside the AI black box. The AI system is not human, of course, but it is not unprecedented for the law to confer legal personhood on other nonhuman entities, including corporations and pet animals. Similarly, the law could confer legal personhood on the AI system so it can be held liable for copyright infringement. This legal fiction would make the AI system the direct copyright infringer.

Holding the AI system liable does not mean that there is no liability for the developer or user, or that copyright owners cannot recover for infringements because the AI system does not have financial resources. Developers or users can be *secondarily* liable for the AI system through their own actions. The law often holds other parties liable for the wrongs of another. For example, employers are responsible for actions their employees carry out in their line of work. Online

conduct. The approach would punish bad actors rather than imposing de facto liability for the mere provision or use of an AI system that also has noninfringing uses. And it has the added benefit of fitting within the historical arc of copyright law remaining flexible enough to adapt to new technologies.

While making the AI system the direct infringer is appropriate under copyright doctrine, this proposal can have broader policy implications for thinking about legal liability in the AI era. AI offers tantalizing benefits, but at the cost of control. To realize the promise of a technology that is valuable precisely because of its increasing autonomy, courts need to consider shifting away from always imposing strict liability on providers or users. Providers can arguably only do so much to prevent harmful outputs ex ante, or before they happen—although tools such as information lattice learning, which attempt to map the black box, could help trace why a particular output occurs. Industry practices, including filtering certain prompts or outputs and reducing the percentage of outputs that reproduces a particular piece of training data, are important practices for countering AI-generated harms. However, a provider cannot predict and evaluate every potential user prompt

The AI system is not human, of course, but it is not unprecedented for the law to confer legal personhood on other nonhuman entities, including corporations and pet animals.

service providers are liable for copyright infringement when they know of a specific infringement and do not remove it.

In the AI context, the developer could be held liable if they knew about and “materially contributed” to the infringement. This could take the form of what I term a *notice-and-revision test*. Under this test, if the developer learns of a specific infringement problem (say, Shane’s lyrics), it would then be obligated to take remedial action to prevent similar infringements from reoccurring. Providers or users could also be on the hook if they induced or encouraged the AI system to infringe. These approaches look to whether the developer or user intended infringement to occur or be furthered, rather than imposing absolute direct liability or a type of principal-agent relationship that would also result in absolute liability for infringements.

AI liability reimagined

By reimagining the AI system as the copyright infringer, courts would not only be faithful to the law but can also have a more nuanced discussion about who else should be held responsible when generative AI infringes. This achieves the purpose of the law by deterring foreseeable, unlawful

and output in advance. This lack of predictability is exacerbated by the iterative nature of machine learning. Shifting the focus of law to enable ex post measures, such as notice-and-revision, may avoid imposing unduly high costs on would-be market entrants while still holding human actors liable when they are ill-intentioned.

Such an equilibrium between absolute and nuanced liability may require creative lawyering as well as judging. My proposal for copyright law would require a novel application of direct liability and a refinement of secondary liability. This allows copyright law to remain flexible and adapt to AI and other emerging technologies, which will continue to evolve and strain the bounds of copyright and other areas of the law. It also puts intent front and center by punishing AI providers and users for engaging in bad-faith actions that facilitate infringement. Together, this strikes a balance between regulating and encouraging new technologies that is ultimately aimed at benefitting society.

Michael P. Goodyear is an acting assistant professor at New York University School of Law and a fellow at NYU’s Engelberg Center on Innovation Law & Policy.