International Center for Law & Economics

Comments of the International Center for Law & Economics to the U.S. Copyright Office, Notice of Inquiry and Request for Comments

Artificial Intelligence and Copyright, Docket No. 2023-6

October 30, 2023

Authors:

Kristian Stout, Director of Innovation Policy
Geoffrey A. Manne, President & Founder
Emily Corbeille, Research Fellow

Comments of the International Center for Law & Economics, Artificial Intelligence and Copyright, U.S. Copyright Office Docket No. 2023-6

Kristian Stout, Geoffrey A. Manne, & Emily Corbeille[*]

Introduction

We thank you for the opportunity to comment on this important notice of inquiry (NOI)[1] on artificial intelligence (AI) and copyright. We appreciate the U.S. Copyright Office undertaking a comprehensive review of the policy and copyright-law issues raised by recent advances in generative AI systems. This NOI covers key areas that require attention, from legal questions regarding infringement and fair use, to questions about how policy choices could shape opportunities for creators and AI producers to engage in licensing.

At this early date, AI systems have already generated some incredible visual art and impressive written texts, as well as a good deal of controversy. Some artists have banded together as part of an anti-AI campaign;[2] lawsuits have been filed;[3] and policy experts have attempted to think through the various legal questions raised by these machine-learning systems.

The debates over the role of AI in creative industries have particular salience for intellectual-property rights. Copyright is notoriously difficult to protect online, and the emergence of AI may exacerbate that difficulty. AI systems also potentially pose an additional wrinkle: it is at least arguable that the outputs they produce can themselves be considered unique creations. There are, of course, other open questions whose answers are relevant here, not the least being whether it is fair to assert that only a human can be "creative" (at least, so far).[4]

But leaving these questions aside, we can say that at least some AI systems produce *unique* outputs and are not merely routinely duplicating other pieces of work in a digital equivalent of collage. That is, at some level, the machines are engaged in a rudimentary sort of "learning" about how humans

---

[*] Kristian Stout is director of innovation policy with the International Center for Law & Economics (ICLE). Geoffrey A. Manne is ICLE's president and founder. Emily Corbeille is an ICLE research fellow.

[1] *Artificial Intelligence and Copyright*, Notice of Inquiry and Request for Comments, U.S. COPYRIGHT OFFICE, LIBRARY OF CONGRESS (Aug. 30, 2023) [*hereinafter* "NOI"].

[2] Tim Sweeney (@TimSweeneyEpic), TWITTER (Jan. 15, 2023, 3:35 AM), https://twitter.com/timsweeneyepic/status/1614541807064608768?s=46&t=0MH_nl5w4PJJl46J2ZT0Dw.

[3] *Pulitzer Prize Winner and Other Authors Accuse OpenAI of Misusing Their Writing*, COMPETITION POLICY INTERNATIONAL (Sep. 11, 2023), https://www.pymnts.com/cpi_posts/pulitzer-prize-winner-and-other-authors-accuse-openai-of-misusing-their-writing; *Getty Images Statement*, GETTY IMAGES (Jan. 17, 2023), https://newsroom.gettyimages.com/en/getty-images/getty-images-statement.

[4] *See, e.g.*, Anton Oleinik, *What Are Neural Networks Not Good At? On Artificial Creativity*, 6 BIG DATA & SOCIETY (2019), *available at* https://journals.sagepub.com/doi/full/10.1177/2053951719839433#bibr75-2053951719839433.

arrange creative inputs when generating images, music, or written works. The machines appear to be able to reconstruct this process and produce new sets of words, sounds, or lines and colors that conform to the patterns found in human art, in at least a simulacrum of "creativity."

But that conclusion isn't the end of the story. Even if some of these AI outputs are unique and noninfringing, the way that AI systems learn—by ingesting massive quantities of existing creative work—raises a number of thorny copyright-law issues. Indeed, some argue that these systems inherently infringe copyright during the learning phase and that, as discussed below, such processes may not survive a "fair use" analysis.

But nor is that assertion the end of the analysis. Rather, it raises the question of whether applying existing doctrine in this novel technological context yields the best results for society. Moreover, it heightens the need for a comprehensive analytical framework to help parse these questions.

A.   The Law & Economics of Copyright and AI

Nearly all would agree that it is crucial that law and public policy strike the appropriate balance between protecting creators' existing rights and enabling society to enjoy the potentially significant benefits that could arise from the development of AI systems. Indeed, the subject is often cast as a dramatic conflict between creative professionals struggling to make ends meet and innovative firms working to provide cutting-edge AI technology. For the moment, however, it is likely more important to determine the right questions to ask and the proper analytical framework to employ than it is to identify any precise balancing point.

What is important to remember is that copyright policy is foremost *economic* in nature and "can be explained as a means for promoting efficient allocation of resources."[5] That is to say, the reason that property rights in creative expression exist is to guarantee the continued production of such works.[6] The fundamental tradeoff in copyright policy is between the costs of limiting access to creative works, and the value obtained by encouraging production of such works.[7] The same applies in the context of AI: identifying the key tradeoffs and weighing the costs and benefits of restricting access to protected works by the producers (and users) of AI systems.[8]

This entails examining the costs and benefits of relatively stronger or weaker forms copyright protection in terms of their effects on both incentives and access, and as they relate to both copyright holders and AI-system developers. It also requires considering where the transaction costs should be allocated for negotiating access to both copyright and, as discussed *infra*,[9] the use of name/image/likeness, as well as how those allocations are likely to shape outcomes.

---

[5] William M. Landes & Richard A. Posner, *An Economic Analysis of Copyright Law*, 18 J. LEGAL STUD. 325 (1989).

[6] *Id.* at 332.

[7] *Id.* at 326.

[8] *Id.*

[9] *See infra*, notes 109-110 and accompanying text.

At root, these questions center on how to think about the property rights that limit access to protected works and, possibly even more importantly, how to assign new property rights governing the ability to control the use of a name/image/likeness. As we know from the work of the late Nobel laureate Ronald Coase, the actual demarcation of rights affects parties' abilities to negotiate superior solutions.[10] The development of nuisance law provides a good example of the problem at hand. When a legal regime provides either strict liability or no-liability rules around pollution, parties have little incentive to minimize harmful conduct:

> The factory that has the absolute right to pollute will, if transaction costs are prohibitive, have no incentives to stop (or reduce) pollution even if the cost of stopping would be much less than the cost of pollution to the homeowners. Conversely, homeowners who have an absolute right to be free from pollution will, if transaction costs are prohibitive, have no incentive to take steps of their own to reduce the effects of pollution even if the cost to them of doing so (perhaps by moving away) is less than the cost to the factory of not polluting or of polluting less.[11]

As Coase observed, this class of problem is best regarded as reciprocal in nature, and the allocation of rights matters in obtaining an efficient outcome. This is necessarily so because, when fully considered, B's ability to restrain A from the pollution-generating activity can itself be conceived of as another kind of harm that B can impose on A. Therefore, the problem requires a balancing of the relative harms generated by both A and B in exercising conflicting claims in a particular context.

When thinking about how to minimize harms—whether from pollution or other activity that generates social costs (which is to say, nearly every activity)—the aim is to decide whether "the gain from preventing the harm is greater than the loss which would be suffered elsewhere as a result of stopping the action which produces the harm."[12] Theoretically, in a world without transaction costs, even assignments of no-liability or strict-liability rules could be bargained around. But we do not live in such a world.[13] Thus, "[i]n a world in which there are costs of rearranging the rights established by the legal system [common law and statutory assignments of liability] are, in effect, making a decision on the economic problem and determining how resources are to be employed."[14]

While pollution rules, unlicensed uses of intellectual property, and a host of other activities subject to legal sanction are not typically framed as resource-allocation decisions, it is undeniable that they do have this character. This is true even where legislation attempts to correct deficiencies in the system. We experience a form of blindness when we focus on correcting what may be rightly perceived as problems in a liability regime. Such analysis tends to concentrate attention on particular deficiencies of the system and to nourish the belief that any measure that removes the deficiency is

---

[10] *See generally* R.H. Coase, *The Problem of Social Cost*, 3 J. L. & ECON. 1, 2 (1960).

[11] RICHARD POSNER, ECONOMIC ANALYSIS OF LAW (Aspen 5th ed 1998) 65, 79.

[12] Coase, *supra* note 10, at 27.

[13] *Id.*

[14] *Id.* at 27.

necessarily desirable. It diverts attention from other changes inevitably associated with the corrective measure—changes that may well produce more harm than the original deficiency.[15]

All of this is to say that one solution to the costs generated by the need for AI systems to process a massive corpus of expensive, copyright-protected material is neither to undermine property rights, nor to make AI impossible, but to think about how new property rights could make the system work. It may be that some entirely different form or allocation of property right would facilitate bargaining between rightsholders and AI creators, optimizing resource allocation in a way the existing doctrinal regime may not be able to.

A number of other questions flow from this insight into the allocative nature of copyright. How would the incentives for human creators change under different copyright rules for AI systems, or in the face of additional rights? And how would access to copyrighted works for AI training change with different rules, and what effects would that access have on AI innovation?

Above all, our goal today should be to properly frame the AI and copyright debate by identifying tradeoffs, quantifying effects (where possible), and asking what rules best serve the overall objectives of the copyright system and the social goal of encouraging AI innovation. The best chance of striking the right balance will come from a rigorous framing of the questions and from the use of economic analysis to try to answer them.

B.   Copyright Law and AI: Moving Forward

As the Copyright Office undertakes this inquiry, it is important to recognize that, regardless of how the immediate legal questions around AI and copyright are resolved, the growing capabilities and adoption of generative AI systems will likely necessitate *some* changes in the long term.

The complex questions surrounding the intersection of AI and copyright law admit reasonable arguments on both sides. But AI is here to stay, regardless, and if copyright law is applied in an unduly restrictive manner that substantially hinders socially beneficial AI innovation, it could provoke a broader public-policy backlash that does more to harm copyright's ability to protect creative works than it does to stanch AI's ability to undermine it. Copyright law risks being perceived as an obstruction to technological progress if it is used preemptively to kill AI in the cradle. Such an outcome could galvanize calls for recalibrating copyright's scope and protections in the name of the public interest.

This illustrates the precarious balancing act that copyright law faces in the wake of rapidly evolving technologies like AI. Aggressive copyright restrictions that curtail AI development could instigate a public-policy counter-reaction before Congress and the courts that ultimately undermines copyright's objectives. The judicious course is to adapt copyright law *cautiously* to enable AI's responsible evolution, while resolutely preserving the incentives for human creativity.

---

[15] *Id.* at 42-43.

In the remainder of this analysis, we offer our perspective on the likely outcomes of the AI-copyright issues raised in this NOI, given the current state of the law. These assessments reflect our perspective formed through the rigorous application of established copyright principles and precedent to the novel technological context of generative AI systems. Reasonable arguments rooted in existing doctrine could be made to support different conclusions. We submit these comments not as definitive predictions or normative preferences, but rather as informed appraisals of how courts may analyze AI under present copyright law, absent legislative intervention.

We appreciate the Copyright Office starting this process to modernize copyright law for the AI age. This inquiry is an important first step, but openness to further evolution will be key to promoting progress in both AI and the arts. We believe an open, evidence-based discussion of these issues will lead to balanced solutions that uphold copyright's constitutionally mandated purpose, while allowing responsible AI innovation for the public benefit.

II. The Training of AI Systems and the Applicability of Fair Use

In the NOI, the Copyright Offices asks: "[u]nder what circumstances would the unauthorized use of copyrighted works to train AI models constitute fair use?"[16]

To answer this question, it would be useful to first briefly walk through a high-level example of how AI systems work, in order to address the most relevant points of contact between AI systems and copyright law.

A. A Brief Technical Description of AI Training

AI-generated content is not a single "thing," but a collection of differing processes, each with different implications for the law. For the purposes of this discussion, we will discuss image generation using "generated adversarial networks" (GANs) and diffusion models. Although different systems and different types of content generation will vary, the basic concepts discussed below are nonetheless useful at a general level.[17]

A GAN is a type of machine-learning model that consists of two parts: a generator and a discriminator.[18] The generator is trained to create new images that look like they come from a particular dataset, while the discriminator is trained to distinguish the generated images from real images in its original dataset.[19] The two parts are trained together in an adversarial manner, with the

---

[16] U.S. Copyright Office, Library of Congress, *supra* note 1, at 14.

[17] For more detailed discussion of GANs and Stable Diffusion see Ian Spektor, *From DALL E to Stable Diffusion: How Do Text-to-image Generation Models Work?*, Tryo Labs Blog (Aug. 31, 2022), https://tryolabs.com/blog/2022/08/31/from-dalle-to-stable-diffusion.

[18] *Id.*

[19] *Id.*

generator trying to produce images that can fool the discriminator and the discriminator trying to correctly identify the generated images.[20]

A diffusion model, by contrast, analyzes the distribution of information in an image, as noise is progressively added to it.[21] This kind of algorithm analyzes characteristics of sample images, like the distribution of colors or lines, in order to understand what counts as an accurate representation of a subject (*i.e.*, what makes a picture of a cat look like a cat, and not like a dog).[22]

For example, in the generation phase, diffusion-based systems start with randomly generated noise, and work backward in "denoising" steps to essentially "see" shapes:

> The sampled noise is predicted so that if we subtract it from the image, we get an image that's closer to the images the model was trained on (not the exact images themselves, but the distribution - the world of pixel arrangements where the sky is usually blue and above the ground, people have two eyes, cats look a certain way – pointy ears and clearly unimpressed).[23]

While it is possible that some implementations might be designed in a way that saves copies of the training images,[24] for at least some systems, once the network is trained using these techniques, it will not need to rely on saved copies of input work in order to produce outputs. The models that are produced during training are, in essence, instructions to a different piece of software about how to start with a prompt from a user, a palette of pure noise, and progressively "discover" signal in that image until some new image emerges.

B.   Fair Use

The creator of some of the most popular AI tools, OpenAI, is not shy about their use of protected works in the training phase of the algorithms. In comments to the U.S. Patent and Trademark Office (PTO), OpenAI noted that:

> Modern AI systems require large amounts of data. For certain tasks, that data is derived from existing publicly accessible "corpora"… of data that include copyrighted works. By analyzing large corpora (**which necessarily involves first making copies of the data to be analyzed**), AI systems can learn patterns inherent in human-generated data and then use

---

[20] *Id.*

[21] *Id.*

[22] *Id.*

[23] Jay Alammar, *The Illustrated Stable Diffusion*, BLOG (Oct. 4, 2022), https://jalammar.github.io/illustrated-stable-diffusion.

[24] Indeed, there is evidence that some models may be trained in a way that they "memorize" their training set, to at least some extent. *See, e.g.,* Kent K. Chang, Mackenzie Cramer, Sandeep Soni, & David Bamman, *Speak, Memory: An Archaeology of Books Known to ChatGPT/GPT-4*, ARXIV PREPRINT (Oct. 20, 2023), https://arxiv.org/abs/2305.00118; OpenAI LP, *Comment Regarding Request for Comments on Intellectual Property Protection for Artificial Intelligence Innovation,* Before the USPTO, DEP'T OF COM. (2019), *available at* https://www.uspto.gov/sites/default/files/documents/OpenAI_RFC-84-FR-58141.pdf.

those patterns to synthesize similar data which yield increasingly compelling novel media in modalities as diverse as text, image, and audio. (emphasis added).[25]

Thus, at the training stage, the most popular forms of AI systems *require* making copies of existing works. And where that material is either not in the public domain or is not licensed, an infringement can occur. Thus, the copy must not be infringing (say, because it is transient), or some affirmative defense is needed to excuse the infringement. Toward this end, OpenAI believes that this use should qualify as fair use,[26] as do most or all the other major producers of generative AI systems.[27]

But as OpenAI has framed the fair-use analysis, it is not clear that these uses should qualify. There are two major questions in this respect: will the data used to train these systems count as "copies" under the Copyright Act, and, if so, is the use of these "copies" sufficiently "transformative" to qualify for the fair-use defense?

1.   Are AI systems being trained with 'copies' of protected works?

Section 106 of the Copyright Act grants the owner of a copyright the exclusive right "to reproduce… copyrighted work in copies" and to authorize others to do so.[28] If an AI system makes a copy of a file to a computer during training, this would likely constitute a *prima facie* violation of the copyright owner's exclusive right of reproduction under Section 106. This is fairly straightforward.

But what if the "copy" is "transient" and/or only partial pieces of content are used in the training? For example, what if a training program merely streamed small bits of a protected work into temporary memory as part of its training, and retained no permanent copy?

As the Copyright Office has previously observed, even temporary reproductions of a work in a computer's memory can constitute "copies" under the Copyright Act.[29] Critically, this includes even temporary reproductions made as part of a packet-switching network transmission, where a particular file is broken into individual packets, because the packets can be reassembled into substantial portions or even entire works.[30] On the topic of network-based transmission, the Copyright Office further observed that:

> Digital networks permit a single disk copy of a work to meet the demands of many users by creating multiple RAM copies. These copies need exist only long enough to be

---

[25] OpenAI, LP, *Comment Regarding Request for Comments on Intellectual Property Protection for Artificial Intelligence*, *id.* (emphasis added).

[26] 17 U.S.C. § 107.

[27] *See, e.g.,* Blake Brittain, *Meta Tells Court AI Software Does Not Violate Author Copyrights*, REUTERS (Sep. 19, 2023), https://www.reuters.com/legal/litigation/meta-tells-court-ai-software-does-not-violate-author-copyrights-2023-09-19; Avram Piltch, *Google Wants AI Scraping to be 'Fair Use.' Will That Fly in Court?*, TOM'S HARDWARE (Aug. 11, 2023), https://www.tomshardware.com/news/google-ai-scraping-as-fair-use.

[28] 17 U.S.C. § 106.

[29] REGISTER OF COPYRIGHTS, DMCA SECTION 104 REPORT (U.S. Copyright Office, Aug. 2001), at 108-22, *available at* https://www.copyright.gov/reports/studies/dmca/sec-104-report-vol-1.pdf.

[30] *Id.* at 122-23.

perceived (e.g., displayed on the screen or played through speakers), reproduced or otherwise communicated (e.g., to a computer's processing unit) in order for their economic value to be realized. If the network is sufficiently reliable, users have no need to retain copies of the material. **Commercial exploitation in a network environment can be said to be based on selling a right to perceive temporary reproductions of works.**[31]

This is a critical insight that translates well to the context of AI training. The "transience" of the copy matters with respect to the receiver's ability to perceive the work in a way that yields commercial value. Under this reasoning, the relevant locus of analysis is on the AI system's ability to "perceive" a work for the purposes of being trained to "understand" the work. In this sense, you could theoretically find the existence of even more temporary copies than that necessary for human perception to implicate the reproduction right.

Even where courts have been skeptical of extending the definition of "copy" to "fleeting" copies in computer memory, this underlying logic is revealed. In *Cartoon Network LP, LLLP v. CSC Holdings, Inc.*, 536 F.3d 121 (2008), the 2nd U.S. Circuit Court of Appeals had to determine whether buffered media sent to a DVR device was too "transient" to count as a "copy":

> No bit of data remains in any buffer for more than a fleeting 1.2 seconds. And unlike the data in cases like MAI Systems, which remained embodied in the computer's RAM memory until the user turned the computer off, each bit of data here is rapidly and automatically overwritten as soon as it is processed. While our inquiry is necessarily fact-specific, and other factors not present here may alter the duration analysis significantly, these facts strongly suggest that the works in this case are embodied in the buffer for only a "transitory" period, thus failing the duration requirement.[32]

In *Cartoon Network*, the court acknowledged both that the duration analysis was fact-bound, and also that the "fleeting" nature of the reproduction was important. "Fleeting" is a relative term, based on the receiver's capacities. A ball flying through the air may look "fleeting" to a human observer, but may appear to go much more cognizable to a creature with faster reaction time, such as a house fly. So, too, with copies of a work in a computer's memory and the ability to "perceive" what is fixed in a buffer: what may be much too quick for a human to perceive may very well be within an AI system's perceptual capabilities.

Therefore, however the training copies are held, there is a strong possibility that a court will find them to be "copies" for the purposes of the reproduction right—even with respect to partial copies that exist for very small amounts of time.

---

[31] *Id.* at 112 (emphasis added).

[32] *Id.* at 129–30.

2.   *The purpose and character of using protected works to train AI systems*

Fair use provides for an affirmative defense against infringement when the use is, among other things, "for purposes such as criticism, comment, news reporting, teaching…, scholarship, or research."[33] When deciding whether a fair-use defense is applicable, a court must balance a number of factors:

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.[34]

The fair-use defense that AI creators have advanced is rooted in the first factor: the nature and character of the use. Although a full analysis of all the factors is ultimately necessary, analysis of the first factor is sufficiently complicated to warrant full attention here. In particular, the complex issue at hand is whether uses of protected works to train AI systems are sufficiently "transformative" or not.[35]

Whether the use of a copyrighted work to train an AI is "transformative" is certainly a novel question, but it is one that will likely be answered in light of an observation the U.S. Supreme Court made in *Campbell v. Acuff Rose Music*:

> [W]hen a commercial use amounts to mere duplication of the entirety of an original, it clearly "supersede[s] the objects,"… of the original and serves as a market replacement for it, making it likely that cognizable market harm to the original will occur… But when, on the contrary, the second use is transformative, market substitution is at least less certain, and market harm may not be so readily inferred.[36]

Moreover, "[t]he word 'transformative' cannot be taken too literally as a sufficient key to understanding the elements of fair use. It is rather a suggestive symbol for a complex thought, and does not mean that any and all changes made to an author's original text will necessarily support a finding of fair use."[37] A key question, then, is whether training AI systems on copyrighted works amounts to a mere "duplication of the entirety of an original" or is sufficiently "transformative" to

---

[33] 17 U.S.C. § 107.

[34] *Id.*; *see also Campbell v. Acuff-Rose Music Inc.*, 510 U.S. 569 (1994).

[35] Critically, a fair use analysis is a multi-factor test, and even within the first factor, it's not a mandatory requirement that a use be "transformative." It is entirely possible that a court balancing all of the factors could indeed find that training AI systems is fair use, even if it does not hold that such uses are "transformative."

[36] Campbell, *supra* note 34, at 591.

[37] *Authors Guild v. Google, Inc.*, 804 F.3d 202, 214 (2d Cir. 2015).

support a fair-use defense. As noted above, OpenAI believes that its use is transformative. According to its comments:

> Training of AI systems is clearly highly transformative. Works in training corpora were meant primarily for human consumption for their standalone entertainment value. The "object of the original creation," in other words, is direct human consumption of the author's expression. Intermediate copying of works in training AI systems is, by contrast, "non-expressive" the copying helps computer programs learn the patterns inherent in human-generated media. The aim of this process—creation of a useful generative AI system—is quite different than the original object of human consumption. The output is different too: nobody looking to read a specific webpage contained in the corpus used to train an AI system can do so by studying the AI system or its outputs. The new purpose and expression are thus both highly transformative.[38]

This framing, however, works against OpenAI's interests. As noted above, and reinforced in the immediately preceding quote, generative AI systems are made of at least two distinct pieces. The first is a piece of software that ingests existing works and creates a file that can serve as instructions to the second piece of software. The second piece of software takes the output of the first and can produce independent results. Thus, there is a clear discontinuity in the process whereby the ultimate work created by the system is disconnected from the creative inputs used to train the software.

Therefore, the protected works are arguably ingested into the first part of the system "for their standalone entertainment value." That is to say, the goal of copying and showing a protected work to an AI system is for the analog of "direct human consumption of the author's expression" in order for the system to learn about that expression.

The software is learning what counts as "standalone entertainment value" and therefore the works *must* be used in those terms. Surely, a computer is not sitting on a couch and surfing for its pleasure. But it is *solely* for the very "standalone entertainment value" that the first piece of software is being shown copyrighted material. By contrast, parody or "remixing" uses incorporate a work into some secondary expression that directly transforms the input. The way these systems work is to learn what makes a piece entertaining and then to discard that piece altogether. Moreover, this use for the art *qua* art most certainly interferes with the existing market, insofar as this use is in lieu of reaching a licensing agreement with rightsholders.

A good analogy is art students and art textbooks. Art students view protected works in an art textbook in order to learn how to reproduce the styles contained therein. The students would not be forgiven for pirating the textbooks merely because they intend to go on to make new paintings. They would still be liable for copyright infringement if they used unlicensed protected works as part of their education.

The 2nd U.S. Circuit Court of Appeals dealt with a case that operates similarly to this dynamic. In *American Geophysical Union v. Texaco*, 60 F.3d 913 (2d Cir. 1994), the 2nd Circuit considered whether

---

[38] OpenAI submission, *supra* note 25, at 5.

Texaco's photocopying of scientific articles produced by the plaintiffs qualified for a fair-use defense. Texaco employed between 400 and 500 research scientists and, as part of supporting their work, maintained subscriptions to a number of scientific journals.[39]

It was common practice for Texaco's scientists to photocopy entire articles and save them in a file.[40] The plaintiffs sued for copyright infringement.[41] Texaco asserted that photocopying by its scientists for the purposes of furthering scientific research—that is *to train the scientists on the content of the journal articles*—should count as a fair use. The argument was, at least in part, that this was sufficiently "transformative," because the scientists were using that knowledge to invent new products.[42] The 2nd Circuit disagreed:

> The "transformative use" concept is pertinent to a court's investigation under the first factor because it assesses the value generated by the secondary use and the means by which such value is generated. **To the extent that the secondary use involves merely an untransformed duplication, the value generated by the secondary use is little or nothing more than the value that inheres in the original.** Rather than making some contribution of new intellectual value and thereby fostering the advancement of the arts and sciences, an untransformed copy is likely to be used simply for the same intrinsic purpose as the original, thereby providing limited justification for a finding of fair use....[43]

The 2nd Circuit thus observed that copies of the scientific articles were made solely to consume the material itself. AI developers often make an argument analogous to that made by Texaco: that training AI systems surely advances scientific research, and therefore fosters the "advancement of the arts and sciences." But in *American Geophysical Union*, the initial copying of copyrighted content, even where it was ultimately used for the "advancement of the arts and sciences," was not held to be sufficiently "transformative."[44] The case thus stands for the proposition that one cannot merely identify a social goal down that would be advanced at some future date in order to permit an exception to copyright protection. As the court put it:

> [T]he dominant purpose of the use is a systematic institutional policy of multiplying the available number of copies of pertinent copyrighted articles by circulating the journals among employed scientists for them to make copies, thereby serving the same purpose for which additional subscriptions are normally sold, or... for which photocopying licenses may be obtained.[45]

---

[39] *Id.* at 915.

[40] *Id.*

[41] *Id.*

[42] *Id.* at 933-34.

[43] *Id.* at 923. (emphasis added)

[44] *Id.*

[45] *Id.* at 924.

The use *itself* must be transformative and different, and copying is not transformative merely because it may be used as an input into a *later* transformative use. By the same token, therefore, it seems likely that where an AI system ingests (copies) copyrighted works, that use is similarly not transformative, despite its ultimate use as an input in the creation of other original works.

Comparing the *American Geophysical Union* analysis with the search-engine "snippets" and "thumbnails" cases provides a useful comparison relevant to the AI analysis. In *Kelly v. Arriba Soft Corp.*, 336 F.3d 811 (9th Cir. 2002), the 9th U.S. Circuit Court of Appeals ruled that a search engine's creation of thumbnail images from original copies was a transformative fair use.[46] Arriba's search-engine crawler made full-sized copies of Kelly's images and stored them temporarily on Arriba's server to generate thumbnail versions. After the thumbnails were created, the full-sized originals were deleted. The thumbnails were used to facilitate Arriba's image-based search engine. In reaching its fair-use conclusion, the 9th Circuit opined that:

> Arriba's use of Kelly's images promotes the goals of the Copyright Act and the fair use exception. The thumbnails do not stifle artistic creativity because they are not used for illustrative or artistic purposes and therefore do not supplant the need for the originals.[47]

Further, although "Arriba made exact replications of Kelly's images, the thumbnails were much smaller, lower-resolution images that served an entirely different function than Kelly's original images."[48]

The court found it important that the search engine did not use the protected works for their intended "aesthetic experience," but rather for the purpose of constructing a search index.[49] Indeed, the entire point of a search engine is not to "supersede" the original, but in many or most cases to provider users an efficient means to find that original online.[50]

The court discussed, but only briefly, the benefit to the public of Arriba's transformative use,[51] noting that "[Arriba's thumbnails] benefit the public by enhancing information-gathering techniques on the internet."[52] Five years later, in *Perfect 10 Inc. v. Amazon.com Inc.*, 487 F.3d 701 (2007), the 9th Circuit expanded on this question somewhat.[53] There, in holding that the novelty of the use was of

---

[46] *Kelly v. Arriba Soft Corp.*, 336 F.3d 811 (9th Cir. 2002).

[47] *Id.*

[48] *Id.* at 818.

[49] *Id.*

[50] *Id.* at 819 ("Arriba's use of the images serves a different function than Kelly's use—improving access to information on the internet versus artistic expression.").

[51] The "public benefit" aspect of copyright law is reflected in the fair-use provision, 17 U.S.C. § 107. In *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 579 (1994), the Supreme Court highlighted the "social benefit" that a use may provide, depending on the first of the statute's four fair-use factors, the "the purpose and character of the use."

[52] *Supra* note 46, at 820.

[53] *Perfect 10 Inc. v. Amazon.com Inc.*, 487 F.3d 701 (9th Cir., 2007)

crucial importance to the analysis,[54] the court also stressed that the *value* of that use was a function of its newness:

> [A] search engine provides social benefit by incorporating an original work into a new work, namely, an electronic reference tool. Indeed, a search engine may be more transformative than a parody [the use at issue in *Campbell*] because a search engine provides an entirely new use for the original work, while a parody typically has the same entertainment purpose as the original work.[55]

Indeed, even in light of the commercial nature of Google's use of copyrighted content in its search engine, its significant public benefit carried the day: "We conclude that the significantly transformative nature of Google's search engine, **particularly in light of its public benefit**, outweighs Google's superseding and commercial uses of the thumbnails in this case."[56] And, of particular relevance to these questions in the context of AI, the court in *Perfect 10* went on to "note the importance of analyzing fair use flexibly **in light of new circumstances**."[57]

Ultimately, the *Perfect 10* decision tracked *Kelly* fairly closely on the rest of the "transformativeness" analysis in finding fair use, because "[a]lthough an image may have been created originally to serve an entertainment, aesthetic, or informative function, a search engine transforms the image into a pointer directing a user to a source of information."[58]

The core throughline in this line of cases is the question of whether a piece of content is being used for its expressive content, weighed against the backdrop of whether the use is for some new (and, thus, presumptively valuable) purpose. In *Perfect 10* and *Kelly*, the transformative use was the creation of a search index.

"Snippets" fair-use cases track a similar line of reasoning. For example, in *Authors Guild v. Google Inc.*, 804 F.3d 202 (2d Cir. 2015), the 2nd Circuit ruled that Google's use of "snippets" of copyrighted books in its Library Project and Google Books website was a "transformative" fair use.[59] Holding that the "snippet view" of books digitized as part of the Google Books project did not constitute an effectively competing substitute to the original works, the circuit court noted that copying for the purpose of "criticism" or—as in that case—copying for the purpose of "provision of information about" the protected work, "tends most clearly to satisfy *Campbell's* notion of the 'transformative' purpose."[60]

---

[54] *Id.* at 721 ("Although an image may have been created originally to serve an entertainment, aesthetic, or informative function, a search engine transforms the image into a pointer directing a user to a source of information.").

[55] *Id.* at 721.

[56] *Id.* at 723 (emphasis added).

[57] *Id.* (emphasis added).

[58] *Id.*

[59] *Supra* note 37, at 218.

[60] *Id.* at 215-16.

Importantly, the court emphasized the importance of the public-benefit aspect of transformative uses: "[T]ransformative uses tend to favor a fair use finding because a transformative use is one that communicates something new and different from the original or expands its utility, **thus serving copyright's overall objective of contributing to public knowledge.**"[61]

Underscoring the idea that the "transformativeness" analysis weighs whether a use is merely for expressive content against the novelty/utility of the intended use, the court observed:

> Google's division of the page into tiny snippets is designed to show the searcher just enough context surrounding the searched term to help her evaluate whether the book falls within the scope of her interest (without revealing so much as to threaten the author's copyright interests). Snippet view thus adds importantly to the highly transformative purpose of identifying books of interest to the searcher.[62]

Thus, the absence of use of the work's expressive content, coupled with a fairly circumscribed (but highly novel) use was critical to the outcome.

The entwined questions of transformative use and the public benefit it confers are significantly more complicated in the AI context, however. Unlike the incidental copying involved in search-engine indexing or thumbnails, training generative AI systems directly leverages copyrighted works for their expressive value. In the *Google Books* and *Kelly* cases, the defendant systems extracted limited portions of works or down-sampled images solely to identify and catalog their location for search purposes. The copies enabled indexing and access, and they expanded public knowledge through a means unrelated to the works' protected aesthetics.

But in training AI models on copyrighted data, the systems necessarily parse the intrinsic creative expression of those works. The AI engages with the protected aesthetic elements themselves, not just superficial markers (like title, length, location on the internet, *etc.*), in order to internalize stylistic and compositional principles. This appropriates the heart of the works' copyright protection for expressive ends, unlike the more tenuous connections in search systems.

The AI is thus "learning" directly from the protected expression in a manner akin to a human student studying an art textbook, or like the scientists learning from the journals in *American Geophysical Union*. The subsequent AI generations are built from mastery of the copyrighted training materials' creative expression. Thus, while search-engine copies only incidentally interact with protected expression to enable unrelated innovation, AI training is predicated on excavating the protected expression itself to fuel iterative creation. These meaningfully different purposes have significant fair-use implications.

---

[61] *Id.* at 214. *See also id.* ("The more the appropriator is using the copied material for new, transformative purposes, the more it serves copyright's goal of enriching public knowledge and the less likely it is that the appropriation will serve as a substitute for the original or its plausible derivatives, shrinking the protected market opportunities of the copyrighted work.").

[62] *Id.* at 218.

This functional difference is, as noted, central to the analysis of a use's "purpose and character." Indeed, "even making an exact copy of a work may be transformative so long as the copy serves a different function than the original work."[63] But the benefit to the public from the new use is important, as well, particularly with respect to the possible legislative response that a restrictive interpretation of existing doctrine may engender.

If existing fair-use principles prohibit the copying required for AI, absent costly item-by-item negotiation and licensing, the transaction costs could become prohibitive, thwarting the development of technologies that promise great public value.[64] Copyright law has faced similar dilemmas before, where the transaction costs of obtaining permission for socially beneficial uses could frustrate those uses entirely.[65] In such cases, we have developed mechanisms like compulsory licensing to facilitate the necessary copying, while still attempting to compensate rightsholders. An unduly narrow fair-use finding for AI training could spur calls for similar interventions in service of enabling AI progress.

In other words, regardless of the veracity of the above conclusion that AI's use of copyrighted works may not, in fact, serve a different function than the original, courts and legislators may be reluctant to allow copyright doctrine to serve as an absolute bar against self-evidently valuable activity like AI development. Our aim should be to interpret or recalibrate copyright law to permit such progress while upholding critical incentives for creators.

C.   Opt-In vs. Opt-Out Use of Protected Works

The question at the heart of the prior discussion—and, indeed, at the heart of the economic analysis of copyright—is whether the transaction costs that accompany requiring express *ex ante* permission for the use of protected works are so high that they impedes socially beneficial conduct whose value would outweigh the social cost of allowing permissionless and/or uncompensated use.[66] The NOI alludes to this question when it asks: "Should copyright owners have to affirmatively consent (opt in) to the use of their works for training materials, or should they be provided with the means to object (opt out)?"[67]

This is a complex problem. Given the foregoing thoughts on fair use, it seems quite possible that, at present, the law requires creators of AI systems to seek licenses for protected content, or else must

---

[63] *Perfect 10*, 487 F.3d at 721-22 (citing *Kelly*, 336 F.3d at 818-19). *See also Campbell*, 510 U.S. at 579 ("The central purpose of this investigation is to see, in Justice Story's words, whether the new work merely 'supersede[s] the objects' of the original creation, or instead adds something new, with a further purpose or different character….") (citations omitted).

[64] *See supra*, notes 10-15 and accompanying text.

[65] *See, e.g.*, the development of the compulsory "mechanical royalty," now embodied in 17 U.S.C. § 115, that was adopted in the early 20th century as a way to make it possible for the manufacturers of player pianos to distribute sheet music playable by their instruments.

[66] *See supra* notes 10-15 and accompanying text.

[67] U.S. COPYRIGHT OFFICE, LIBRARY OF CONGRESS, *supra* note 1, at 15.

resort to public-domain works for training. Given the volume of copyrighted works that AI developers currently use to train these systems, such requirements may be broadly infeasible.

On one hand, requiring affirmative opt-in consent from copyright holders imposes significant transaction costs on AI-system developers to identify and negotiate licenses for the vast amounts of training data required. This could hamper innovation in socially beneficial AI systems. On the other hand, an opt-out approach shifts more of the transaction-cost burden to copyright holders, who must monitor and object to unwanted uses of their works. This raises concerns about uncompensated use.

Ultimately, the question is where the burden should lie: with AI-system developers to obtain express consent, or with copyright holders to monitor and object to uses? Requiring some form of consent may be necessary to respect copyright interests. Yet an opt-out approach may strike the right balance, by shifting some of the burden back to AI developers while avoiding the infeasibly high transaction costs of mandatory opt-in consent. The optimal approach likely involves nuanced policymaking to balance these competing considerations. Moreover, as we discuss *infra*, the realistic outcome is most likely going to require rethinking the allocation of property rights in ways that provide for large-scale licensing. Ideally, this could be done through collective negotiation, but perhaps at a *de minimis* rate, while allowing creators to bargain for remuneration on the basis of other rights, like a right of publicity or other rights attached to the output of AI systems, rather than the inputs.[68]

1.   Creator consent

Relatedly, the Copyright Office asks: "If copyright owners' consent is required to train generative AI models, how can or should licenses be obtained?"[69]

Licensing markets exist, and it is entirely possible that major AI developers and large groups of rightsholders can come to mutually beneficial terms that permit a sufficiently large body of protected works to be made available as training data. Something like a licensing agency for creators who choose to make their works available could arise, similar to the services that exist to provide licensed music and footage for video creators.[70] It is also possible for some to form collective-licensing organizations to negotiate blanket permissions covering many works.

It's important to remember that our current thinking is constrained by our past experience. All we know today are AI models trained on vast amounts of unlicensed works. It is entirely possible that, if firms were required to seek licenses, unexpected business models would emerge to satisfy both sides of the equation.

For example, an AI firm could develop its own version of YouTube's ContentID, which would allow creators to control when their work is used in AI training. For some well-known artists, this could

---

[68] *See infra*, notes at 109-110 and accompanying text.

[69] U.S. COPYRIGHT OFFICE, LIBRARY OF CONGRESS, *supra* note 1, at 15.

[70] *See, e.g.*, Copyright Free Music, Premium Beat By Shutterstock, https://www.premiumbeat.com/royalty-free/licensed-music; Royalty-free stock footage at your fingertips, Adobe Stock, https://stock.adobe.com/video.

be negotiated with an upfront licensing fee. On the user side, any artist who has opted in could then be selected as a "style" for the AI to emulate—triggering a royalty payment to the artist when a user generates an image or song in that style. Creators could also have the option of removing their influence from the system if they so desire.

Undoubtedly, there are other ways to structure the relationship between creators and AI systems that would facilitate creators' monetization of the use of their work in AI systems, including legal and commercial structures that create opportunities for both creators and AI firms to succeed.

III.   Generative AI Outputs: Protection of Outputs and Outputs that Infringe

The Copyright Office asks: "Under copyright law, are there circumstances when a human using a generative AI system should be considered the 'author' of material produced by the system?"[71]

Generally speaking, we see no reason why copyright law should be altered to afford protection to purely automatic creations generated by AI systems. That said, when a human makes a nontrivial contribution to generative AI output—such as editing, reframing, or embedding the AI-generated component within a larger work—the resulting work should qualify for copyright protection.

Copyright law centers on the concept of original human authorship.[72] The U.S. Constitution expressly limits copyright to "authors."[73] As of this writing, however, generative AI's capacities do not rise to the level of true independent authorship. AI systems remain tools that require human direction and judgment.[74] As such, when a person provides the initial prompt or framing, makes choices regarding the iterative development of the AI output, and decides that the result is satisfactory for inclusion in a final work, they are fundamentally engaging in creative decision making that constitutes authorship under copyright law.

As Joshua Gans has observed of recent Copyright Review Board decisions:

> Trying to draw some line between AI and humans with the current technology opens up a massive can of worms. There is literally no piece of digital work these days that does not have some AI element to it, and some of these mix and blur the lines in terms of what is creative and what is not. Here are some examples:
>
> A music artist uses AI to denoise a track or to add an instrument or beat to a track or to just get a composition started.

---

[71] U.S. Copyright Office, Library of Congress, *supra* note 1, at 19.

[72] *Id.*

[73] U.S. Const. art. I, § 8, cl. 8.

[74] *See* Ajay Agrawal, Joshua S. Gans, & Avi Goldfarb, *Exploring the Impact of Artificial Intelligence: Prediction Versus Judgment*, 47 Info. Econ. & Pol'y 1, 1 (2019) ("We term this process of understanding payoffs, 'judgment'. At the moment, it is uniquely human as no machine can form those payoffs.").

A photographer uses Photoshop or takes pictures with an iPhone that already uses AI to focus the image and to sort a burst of images into one that is appropriate.

A writer uses AI to prompt for some dialogue when stuck at some point or to suggest a frame for writing a story.[75]

Attempting to separate out an "AI portion" from the final work, as the Copyright Review Board proposed, fundamentally misunderstands the integrated nature of the human-AI collaborative process. The AI system cannot function without human input, and its output remains raw material requiring human creativity to incorporate meaningfully into a finished product.

Therefore, when a generative AI system is used as part of a process guided by human creative choices, the final work should be protected by copyright, just as a work created using any other artistic tool or collaborator would be. Attenuating copyrightability due to the use of AI would undermine basic copyright principles and fail to recognize the essentially human nature of the creative process.

A.   AI Outputs and Infringement

The NOI asks: "Is the substantial similarity test adequate to address claims of infringement based on outputs from a generative AI system, or is some other standard appropriate or necessary?" (Question 23)

The outputs of AI systems may or may not violate IP laws, but there is nothing inherent in the processes described above that dictates that they must. As noted, the most common AI systems do not save copies of existing works, but merely "instructions" (more or less) on how to create new work that conforms to patterns found by examining existing work. If we assume that a system isn't violating copyright at the input stage, it's entirely possible that it can produce completely new pieces of art that have never before existed and do not violate copyright.

They can, however, be made to violate copyrights. For example, these systems can be instructed to generate art, not just in the style of a particular artist, but art that very closely resembles existing pieces. In this sense, it would be making a copy that theoretically infringes. The fact of an AI's involvement would not change the analysis: just as with a human-created work, if it is substantially similar to a copyrighted work, it may be found infringing.

There is, however, a common bug in AI systems that leads to outputs that are more likely to violate copyright in this way. Known as "overfitting," the training leg of these AI systems can be presented with samples that contain too many instances of a particular image.[76] This leads to a dataset that contains too much information about the specific image, such that—when the AI generates a new image—it is constrained to producing something very close to the original. Similarly, there is evidence

---

[75] Joshua Gans, *Can AI works get copyright protection? (Redux)*, Joshua Gans' Newsletter (Sept. 7, 2023), https://joshuagans.substack.com/p/can-ai-works-get-copyright-protection.

[76] *See* Nicholas Carlini, *et al.*, *Extracting Training Data from Diffusion Models*, Cornell Univ. (Jan. 30, 2023), *available at* https://arxiv.org/abs/2301.13188.

that some AI systems are "memorizing" parts of protected books.[77] This could lead to AI systems repeating copyright-protected written works.

1.   The substantial-similarity test

The substantial-similarity test remains functionally the same when evaluating works generated using AI. To find "substantial similarity," courts require evidence of copying, as well as an expression that is substantially similar to a protected work.[78] "It is now an axiom of copyright law that actionable copying can be inferred from the defendant's access to the copyrighted work and substantial similarity between the copyrighted work and the alleged infringement."[79] In many or most cases, it will arguably be the case that AI systems have access to quite a wide array of protected works that are posted online. Thus, there may not be a particularly high hurdle to determine that an AI system actually copied a protected work.

There is, however, one potential problem for the first prong of this analysis. Models produced during a system's training process do not (usually) contain the original work, but are the "ideas" that the AI systems generated during training. Thus, where the provenance of works contained in a training corpus is difficult to source, it may not be so straightforward to make inferences about whether a model "saw" a particular work. This is because the "ideas" that the AI "learns" from its training corpus are unprotected under U.S. copyright law, as it is permissible to mimic unprotected elements of a copyrighted work (such as ideas).[80]

Imagine a generative AI system trained on horror fiction. It would be possible for this system to produce a new short story that is similar to one written by Stephen King, but the latent data in the model almost certainly would not violate any copyrights that King holds in his work. The model would contain "ideas" about horror stories, including those learned from an array of authors who were themselves influences on Stephen King, and potentially some of King's own stories. What the AI system "learns" in this case is the relationship between words and other linguistic particularities that are commonly contained in horror fiction. That is, it has "ideas" about what goes into a horror story, not (theoretically) the text of the horror story itself.

Thus, when demonstrating indirect proof of copying in the case of a Stephen King story, it may pose a difficulty that an AI system has ingested all of H.P. Lovecraft's work—an author who had a major influence on King. The "ideas" in the model and the output it subsequently produces may, in fact, produce something similar to a Stephen King work, but it may have been constructed largely or

---

[77] *See* Chang, Cramer, Soni, & Bamman, *supra* note 24; *see also* Matthew Sag, *Copyright Safety for Generative AI*, Working Paper (May 4, 2023), *available at* https://ssrn.com/abstract=4438593.; Andrés Guadamuz, *A Scanner Darkly: Copyright Liability and Exceptions in Artificial Intelligence Inputs and Outputs*, 25-27 (Mar. 1, 2023), *available at* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4371204.

[78] *Laureyssens v. Idea Grp. Inc.*, 964 F.2d 131, 140 (2d Cir. 1992), as amended (June 24, 1992).

[79] *Id.* at 139.

[80] *Harney v. Sony Pictures Television Inc.*, 704 F.3d 173, 178 (1st Cir. 2013). This assumes, for argument's sake, that a given model is not "memorizing," as noted above.

entirely on material from Lovecraft and other public-domain horror writers. The problem becomes only more complicated when you realize that this system could also have been trained on public-domain fan fiction written in the style of Stephen King. Thus, for the purposes of the first prong of this analysis, courts may place greater burden on plaintiffs in copyright actions against model producers to demonstrate more than merely that a work was merely available online.

Assuming that plaintiffs are able to satisfy the first prong, once an AI system "expresses" those ideas, that expression could violate copyright law under the second prong of the substantial-similarity test. The second prong inquires whether the final work appropriated the protected original expression.[81] Any similarities in unprotectable ideas, facts, or common tropes are disregarded.[82] So, in both traditional and AI contexts, the substantial-similarity test ultimately focuses on the protected components of creative expression, not surface similarity.

The key determination is whether the original work's protected expression itself has been impermissibly copied, no matter the process that generated the copy. AI is properly viewed as simply another potential tool that could be used in certain acts of copying. It does not require revisiting settled principles of copyright law.

B.   Direct and Secondary Liability

The NOI asks: "If AI-generated material is found to infringe a copyrighted work, who should be directly or secondarily liable—the developer of a generative AI model, the developer of the system incorporating that model, end users of the system, or other parties?"[83]

Applying traditional copyright-infringement frameworks to AI-generated works poses unique challenges in determining direct versus secondary liability. In some cases, the AI system itself may create infringing content without any direct human causation.

1.   Direct liability

If the end user prompts an AI system in a way that intentionally targets copyrighted source material, they may meet the threshold for direct infringement by causing the AI to reproduce protected expression.[84] Though many AI prompts contain only unprotected ideas, users may sometimes input copyrightable material as the basis for the AI output. For example, a user could upload a copyrighted image and request the AI to make a new drawing based on the sample. In such cases, the user is

---

[81] *Id.* at 178-79.

[82] *Id.*

[83] U.S. Copyright Office, Library of Congress, *supra* note 1, at 25.

[84] Notably, the state of mind of the user would be irrelevant from the point of view of whether an infringement occurs. All that is required is that a plaintiff owns a valid copyright, and that the defendant infringed it. 17 U.S.C. 106. There are cases where the state of mind of the defendant will matter, however. For one, willful or recklessly indifferent infringement by a plaintiff will open the door for higher statutory damages. *See, e.g.*, Island Software & Computer Serv., Inc. v. Microsoft Corp., 413 F.3d 257, 263 (2d Cir. 2005). For another, a case of criminal copyright infringement will require that a defendant have acted "willfully." 17 U.S.C. § 506(a)(1) (2023), 18 U.S.C. § 2319 (2023).

intentionally targeting copyrighted works and directly "causing" the AI system to reproduce output that is similar. If sufficiently similar, that output could infringe on the protected input. This would be a question of first impression, but it is a plausible reading of available cases.

For example, in *CoStar Grp. Inc. v. LoopNet Inc.*, 373 F.3d 544 (4th Cir. 2004), the 4th U.S. Circuit Court of Appeals had to consider whether an internet service provider (ISP) could be directly liable when third parties reposted copyrighted material owned by the plaintiff. In determining that merely owning the "machine" through which copies were made or transmitted was not enough to "cause" a direct infringement, the court held that:

> [T]o establish direct liability under §§ 501 and 106 of the Act, something more must be shown than mere ownership of a machine used by others to make illegal copies. **There must be actual infringing conduct with a nexus sufficiently close and causal to the illegal copying that one could conclude that the machine owner himself trespassed on the exclusive domain of the copyright owner. The Netcom court described this nexus as requiring some aspect of volition or causation…** Indeed, counsel for both parties agreed at oral argument that a copy machine owner who makes the machine available to the public to use for copying is not, without more, strictly liable under § 106 for illegal copying by a customer. The ISP in this case is an analogue to the owner of a traditional copying machine whose customers pay a fixed amount per copy and operate the machine themselves to make copies. When a customer duplicates an infringing work, the owner of the copy machine is not considered a direct infringer. Similarly, an ISP who owns an electronic facility that responds automatically to users' input is not a direct infringer.[85]

Implied in the 4th Circuit's analogy is that, while the owner of a copying machine might not be a direct infringer, a user employing such a machine could be a direct infringer. It's an imperfect analogy, but a user of an AI system prompting it to create a "substantially similar" reproduction of a protected work could very well be a direct infringer under this framing. Nevertheless, the analogy is inexact, because the user feeds an original into a copying machine in order to make a more-or-less perfect copy of the original, whereas an AI system generates something new but similar. The basic mechanism of using a machine to try to reproduce a protected work, however, remains essentially the same. Whether there is an infringement would be a question of "substantial similarity."

2. Secondary liability

As in the case of direct liability, the nature of generative AI makes the secondary-liability determination slightly more complicated, as well. That is, paradoxically, the basis for secondary liability could theoretically arise even where there was no direct infringement.[86]

The first piece of this analysis is relatively easier. If a user is directly liable for infringing a protected work, as noted above, the developer and provider of a generative AI system may face secondary

---

[85] *Id.* at 550.

[86] Legally speaking, it would be incoherent to suggest that there can be secondary liability without primary liability. The way that AI systems work, however, could prompt Congress to modify the law in order to account for the identified situation.

copyright liability. If the AI developer or distributor knows the system can produce infringing outputs, and provides tools or material support that allows users to infringe, it may be liable for contributory infringement.[87] Critically, merely designing a system that is *capable* of infringing is not enough to find contributory liability.[88]

An AI producer or distributor may also have vicarious liability, insofar as it has the right and ability to supervise users' activity and a direct financial interest in that activity.[89] AI producers have already demonstrated their ability to control users' behavior to thwart unwanted uses of the service.[90] Thus, if there is a direct infringement by a user, a plausible claim for vicarious liability could be made so long as there is sufficient connection between the user's behavior and the producer's financial interests.

The question becomes more complicated when a user did not direct the AI system to infringe. When the AI generates infringing content without user direction, it's not immediately clear who would be liable for the infringement.[91] Consider the case where, unprompted by either the user or the AI producer, an AI system creates an output that would infringe under the substantial-similarity test. Assuming that the model has not been directed by the producer to "memorize" the works it ingests, the model itself consists of statistical information about the relationship between different kinds of data. The infringer, in a literal sense, is the AI system itself, as it is the creator of the offending output. Technically, this may be a case of vicarious liability, even without an independent human agent causing the direct infringement.

We know that copyright **protection** can only be granted to humans. As the Copyright Review Board recently found in a case deciding whether AI-generated outputs can be copyrighted:

> The Copyright Act protects, and the Office registers, "original works of authorship fixed in any tangible medium of expression." 17 U.S.C. § 102(a). Courts have interpreted the statutory phrase "works of authorship" to require human creation of the work.[92]

---

[87] *See, e.g., Metro-Goldwyn-Mayer Studios Inc. v. Grokster Ltd.*, 380 F.3d 1154, 1160 (9th Cir. 2004), vacated and remanded, 545 U.S. 913, 125 S. Ct. 2764, 162 L. Ed. 2d 781 (2005).

[88] *See BMG Rts. Mgmt. (US) LLC v. Cox Commc'ns Inc.*, 881 F.3d 293, 306 (4th Cir. 2018); *Sony Corp. of Am. v. Universal City Studios Inc.*, 464 U.S. 417, 442 (1984).

[89] *A&M Recs. Inc. v. Napster Inc.*, 239 F.3d 1004, 1022 (9th Cir. 2001), as amended (Apr. 3, 2001), aff'd sub nom. *A&M Recs. Inc. v. Napster Inc.*, 284 F.3d 1091 (9th Cir. 2002), and aff'd sub nom. *A&M Recs. Inc. v. Napster Inc.*, 284 F.3d 1091 (9th Cir. 2002).

[90] *See, e.g., Content Filtering*, MICROSOFT IGNITE, available at https://learn.microsoft.com/en-us/azure/ai-services/openai/concepts/content-filter (last visited Oct. 27, 2023).

[91] Note that, if an AI producer can demonstrate that they used no protected works in the training phase, there may in fact be no liability for infringement at all. If a protected work is never made available to the AI system, even an output very similar to that protected work might not be "substantially similar" in a legal sense.

[92] Copyright Review Board, *Second Request for Reconsideration for Refusal to Register Théâtre D'opéra Spatial (SR # 1-11743923581; Correspondence ID: 1-5T5320R)*, U.S. COPYRIGHT OFFICE (Sep. 5, 2023), *available at* https://fingfx.thomsonreuters.com/gfx/legaldocs/byprrqkqxpe/AI%20COPYRIGHT%20REGISTRATION%20decision.pdf.

But can an AI system directly violate copyright? In his *Aereo* dissent, Justice Clarence Thomas asserted that it was a longstanding feature of copyright law that violation of the performance right required volitional behavior.[93] But the majority disagreed with him, holding that, by running a fully automated system of antennas intended to allow users to view video at home, the system gave rise to direct copyright liability.[94] Thus, implied in the majority's opinion is the idea that direct copyright infringement does *not* require "volitional" conduct.

It is therefore plausible that a non-sentient, fully automated AI system *could* infringe copyright, even if, ultimately, there is no way to recover against the nonhuman agent. That does, however, provide an opportunity for claims of vicarious liability against the AI producer or distributor— at least, where the producer has the power to control the AI system's behavior and that behavior appears to align with the producer's financial interests.

3.   *Protecting the 'style' of human creators*

The NOI asks: "Are there or should there be protections against an AI system generating outputs that imitate the artistic style of a human creator (such as an AI system producing visual works 'in the style of' a specific artist)?"[95]

At the federal level, one candidate for protection against AI imitating some aspects of a creator's works can currently be found in trademark law. Trademark law, governed by the Lanham Act, protects names, symbols, and other source identifiers that distinguish goods and services in commerce.[96] Unfortunately, a photograph or likeness, on its own, typically does not qualify for trademark protection, unless it is consistently used on specific goods.[97] Even where there is a likeness (or similar "mark") used consistently as part of branding a distinct product, many trademark-infringement claims would be difficult to establish in this context, because trademark law does little to protect many aspects of a creator's work.

Moreover, the Supreme Court has been wary about creating a sort of "mutant copyright" in cases that invoke the Lanham Act as a means to enforce a sort of "right of attribution," which would potentially give creators the ability to control the use of their name in broader contexts.[98] In this context, the Court has held that the relevant parts of the Lanham Act were not designed to "protect

---

[93] *Am. Broad. Companies Inc. v. Aereo Inc.*, 573 U.S. 431, 453 (2014). (Thomas J, dissenting).

[94] *Id.* at 451.

[95] U.S. COPYRIGHT OFFICE, LIBRARY OF CONGRESS, *supra* note 1, at 21.

[96] *See* 5 U.S.C. § 1051 et seq. at § 1127.

[97] *See, e.g.*, *ETW Corp. v. Jireh Pub. Inc.*, 332 F.3d 915, 923 (6th Cir. 2003).

[98] *Dastar Corp. v. Twentieth Century Fox Film Corp.*, 539 U.S. 23, 34 (2003).

originality or creativity,"[99] but are focused solely on "actions like trademark infringement that deceive consumers and impair a producer's goodwill."[100]

In many ways, there is a parallel here to the trademark cases involving keyword bidding in online ads. At a high level, search engines and other digital-advertising services do not generally infringe trademark when they allow businesses to purchase ads triggered by a user's search for competitor trademarks (*i.e.*, rivals' business names).[101] But in some contexts, this can be infringing—*e.g.*, where the use of trademarked terms in combination with advertising text can mislead consumers about the origin of a good or service.[102]

Thus, the harm, when it arises, would not be in a user asking an AI system to generate something "in the style of" a known creator, but when that user subsequently seeks to release a new AI-generated work and falsely claims it originated from the creator, or leaves the matter ambiguous and misleading to consumers.

Alternative remedies for creators could be found in the "right of publicity" laws in various states. A state-level right of publicity "is not merely a legal right of the 'celebrity,' but is a right inherent to everyone to control the commercial use of identity and persona and recover in court damages and the commercial value of an unpermitted taking."[103] Such rights are recognized under state common law and statutes, which vary considerably in scope across jurisdictions—frequently as part of other privacy statutes.[104] For example, some states only protect an individual's name, likeness, or voice, while others also cover distinctive appearances, gestures, and mannerisms.[105] The protections afforded for right-of-publicity claims vary significantly based on the state where the unauthorized use occurs or the individual is domiciled.[106] This creates challenges for the application of uniform nationwide protection of creators' interests in the various aspects that such laws protect.

In recent hearings before the U.S. Senate Judiciary Subcommittee on Intellectual Property, several witnesses advocated creating a federal version of the right of publicity.[107] The Copyright Office has

---

[99] *Id.* at 37.

[100] *Id.* at 32.

[101] *See, e.g.*, *Multi Time Mach. Inc. v. Amazon.com Inc.*, 804 F.3d 930, 938 (9th Cir. 2015); *EarthCam Inc. v. OxBlue Corp.*, 49 F. Supp. 3d 1210, 1241 (N.D. Ga. 2014); *Coll. Network Inc. v. Moore Educ. Publishers Inc.*, 378 F. App'x 403, 414 (5th Cir. 2010).

[102] *Digby Adler Grp. LLC v. Image Rent a Car Inc.*, 79 F. Supp. 3d 1095, 1102 (N.D. Cal. 2015).

[103] J. Thomas McCarthy, *The Rights of Publicity and Privacy § 1:3. Introduction–Definition and History of the Right of Publicity–Simple Definition of the Right of Publicity*, 1 RIGHTS OF PUBLICITY AND PRIVACY § 1:3 (2d ed).

[104] *See id.* at § 6:3.

[105] Compare Ind. Code § 32-36-1-7 (covering name, voice, signature, photograph, image, likeness, distinctive appearance, gesture, or mannerism), *with* Ky. Rev. Stat. Ann. § 391.170 (limited to name and likeness for "public figures").

[106] *See* Restatement (Third) of Unfair Competition § 46 (1995).

[107] *See, e.g.*, Jeff Harleston, *Artificial Intelligence and Intellectual Property – Part II: Copyright*, U.S. SENATE COMM. ON THE JUDICIARY SUBCOMM. ON INTELLECTUAL PROPERTY (Jul.12, 2023), *available at* https://www.judiciary.senate.gov/imo/media/doc/2023-07-12_pm_-_testimony_-_harleston1.pdf; Karla Ortiz, *"AI and*

also previously opined that it may be desirable for Congress to enact some form of a "right of publicity" law.[108] If Congress chose to enact a federal "right of privacy" statute, several key issues would need to be addressed regarding the scope of protection, effect on state laws, constitutional authority, and First Amendment limitations.

Congress would have to delineate the contours of the federal right of publicity, including the aspects of identity covered and the types of uses prohibited. A broad right of privacy could protect names, images, likenesses, voices, gestures, distinctive appearances, and biographical information from any unauthorized commercial use. Or Congress could take a narrower approach focused only on particular identity attributes, like name and likeness. Congress would also need to determine whether a federal right-of-publicity statute preempts state right-of-publicity laws or sets a floor that would allow state protections to exceed the federal standards.

4.   Bargaining for the use of likenesses

A federal right of publicity could present an interesting way out of the current dispute between rightsholders and AI producers. Most of the foregoing comment attempts to pull apart different pieces of potential infringement actions, but such actions are only necessary, obviously, if a mutually beneficial agreement cannot be struck between creators and AI producers. The main issue at hand is that, given the vast amount of content necessary to train an AI system, it could be financially impractical for even the largest AI firms to license all the necessary content. Even if the comments above are correct, and fair use is not available, it could very well be the case that AI producers will not license very much content, possibly relying on public-domain material, and choosing to license only a very small selection.

Something like a "right of publicity," or an equivalent agreement between creators and AI producers, could provide alternative licensing and monetization strategies that encourage cooperation between the parties. If creators had the opportunity to opt into the use of their likeness (or the relevant equivalent for the sort of AI system in question), the creators could generate revenue when the AI system actually uses the results of processing their content. Thus, the producers would not need to license content that contributes an unknown and possibly *de minimis* value to their systems, and would only need to pay for individual instances of use.

Indeed, in this respect, we are already beginning to see some experimentation with business models. The licensing of celebrity likenesses for Meta's new AI chatbots highlights an emerging opportunity for creators to monetize their brand through contractual agreements that grant usage rights to tech

---

Copyright", U.S. Senate Comm. on the Judiciary Subcomm. on Intellectual Property (Jul. 7, 2023), *available at* https://www.judiciary.senate.gov/imo/media/doc/2023-07-12_pm_-_testimony_-_ortiz.pdf; Matthew Sag, *"Artificial Intelligence and Intellectual Property – Part II: Copyright and Artificial Intelligence"*, U.S. Senate Comm. on the Judiciary Subcomm. on Intellectual Property (Jul. 12, 2023), *available at* https://www.judiciary.senate.gov/imo/media/doc/2023-07-12_pm_-_testimony_-_sag.pdf.

[108] Authors, Attribution, and Integrity: Examining Moral Rights in the United States, U.S. Copyright Office (Apr. 2019) at 117-119, https://www.copyright.gov/policy/moralrights/full-report.pdf.

companies that commercialize conversational AI.[109] As this technology matures, there will be more opportunities for collaborations between AI producers—who are eager to leverage reputable and recognizable personalities—and celebrities or influencers seeking new income streams.

As noted, much of the opportunity for creators and AI producers to reach these agreements will depend on how rights are assigned.[110] It may be the case that a "right of publicity" is not necessary to make this sort of bargaining happen, as creators could—at least theoretically—pursue litigation on a state-by-state basis. This disparate-litigation strategy could deter many creators, however, and it could also be the case that a single federal standard outlining a minimal property right in "publicity" could help to facilitate bargaining.

Conclusion

The advent of generative AI systems presents complex new public-policy challenges centered on the intersection of technology and copyright law. As the Copyright Office's inquiry recognizes, there are open questions around the legal status of AI-training data, the attribution of AI outputs, and infringement liability, which all require thoughtful analysis.

Ultimately, maintaining incentives for human creativity, while also allowing AI systems to flourish, will require compromise and cooperation between stakeholders. Rather than an outright ban on the unauthorized use of copyrighted works for training data, a licensing market that enables access to a large corpora could emerge. Rightsholders may need to accept changes to how they typically license content. In exchange, AI producers will have to consider how they can share the benefit of their use of protected works with creators.

Copyright law retains flexibility to adapt to new technologies, as past reforms reacting to photography, sound recordings, software, and the internet all demonstrate. With careful balancing of interests, appropriate limitations, and respect for constitutional bounds, copyright can continue to promote the progress of science and the useful arts even in the age of artificial intelligence. This inquiry marks a constructive starting point, although ongoing reassessment will likely be needed as generative AI capabilities continue to advance rapidly.

---

[109] Benj Edwards, *Meta Launches Consumer AI Chatbots with Celebrity Avatars in its Social Apps*, ARSTECHNICA (Sep. 28, 2023), https://arstechnica.com/information-technology/2023/09/meta-launches-consumer-ai-chatbots-with-celebrity-avatars-in-its-social-apps; Max Chafkin, *Meta's New AI Buddies Aren't Great Conversationalists*, BLOOMBERG (Oct. 17, 2023), https://www.bloomberg.com/news/newsletters/2023-10-17/meta-s-celebrity-ai-chatbots-on-facebook-instagram-are-surreal.

[110] *See supra*, notes 9-15 and accompanying text.