

III. WHICH WAY FROM HERE?

The generative-AI supply chain is extremely complex. So is copyright law. Putting the two of them together multiplies the intricacy. Two unsettling conclusions follow from this radiating complexity.

First, because of the complexity of the *supply chain*, it is not possible to make accurate sweeping statements about the copyright legality of generative AI. Too much depends on the details of the specific system in question. All the pieces matter, from the curatorial choices in the training dataset, to the training algorithm, to the deployment environment, to the prompt supplied by the user. Courts will inevitably have to work through these details in numerous lawsuits, as they develop doctrines to distinguish among different systems and uses.

Second, because of the complexity of *copyright law*, there is enormous play in the joints. In particular, substantial similarity, indirect infringement, fair use, and remedies all have open-ended tests that can reach different results depending on the facts a court emphasizes and the conclusions it draws. This complexity gives courts the flexibility to deal with the many variations in the supply chain. Paradoxically, it also gives courts the freedom to reach any of several different plausible conclusions about a generative-AI system.

In this Part, we explore some of the ways that courts might try to use their discretion to apply copyright law to generative AI,⁶¹⁶ and then discuss some of the considerations that courts should keep in mind as they do.⁶¹⁷

A. Possible Outcomes

Although the details of which generative-AI systems fall into which boxes may vary, there are a few boxes that courts may find it appealing to sort them into. In this section, we sketch a few of the possible copyright regimes that might result.

No Liability

First, courts might settle on a regime of no liability for services and their users. Anything produced by a generative-AI system would be categorically legal, under a combination of no substantial similarity and fair use. The result would be that models and services would also be categorically legal — there would be no primary liability for them to be indirectly liable for, and

⁶¹⁶. See *infra* Part III.A.

⁶¹⁷. See *infra* Part III.B.

intermediate nonexpressive fair use would shield them in any event. Training datasets would also usually be legal as well (except perhaps in cases of blatant infringement like Books3).⁶¹⁸ They would be fair -use inputs to non-infringing downstream stages of the supply chain.

This regime is clear and simple. It would also be unstable. While such an outcome might make sense for some generative-AI systems, it seems both unworkable and undesirable for others, including systems trained specifically to emulate the styles of particular creators, and retrieval systems that find matching works and reproduce them nearly exactly.⁶¹⁹ If all generative AI were categorically legal, then developers would plausibly start adding generative components to other systems in order to launder copyrighted works through them. The endpoint could be the effective collapse of copyright. On the assumption that this is not an outcome that courts would willingly preside over, then, a blanket no-liability regime seems unlikely. Instead, courts would be more likely to find at least some infringement — so the question becomes where to draw the line.

Liability for Generations Only

Second, courts could draw a line between generative-AI services and the users of those services. In this regime, only generations would be treated as infringing, and then only when a user made some external use of them.⁶²⁰ In this world, generative-AI systems would be creative tools like Photoshop.⁶²¹ The user would be responsible for making sure that anything they create with the tools is noninfringing, but the tools would be shielded under something like a strong *Sony* rule, assembled out of a combination of no substantial similarity, no indirect infringement, and/or fair use. This result might be unfair to users whose infringements resulted from systems producing generations that reproduce material in the underlying model's training dataset, through no choice or fault of their own. But this is arguably the same kind of situation that some courts currently countenance when they hold that users can be liable for embedding images from Instagram even though Instagram is

618. Knibbs, *supra* note 521; Reisner, *supra* note 521; Complaint, *Kadrey v. Meta Platforms, Inc.*, No. 3:23-cv-03417 (N.D. Cal. July 7, 2023).

619. See *supra* note 535 and accompanying text.

620. Here, we use the term “user” broadly. A user could be a customer using a web application to produce a generation, a developer using an API to produce a generation in their own code, a developer using an API to produce a generation for a company, etc.

621. Sometimes literally so. See Adobe, *Experience the Future of Photoshop With Generative Fill* (July 27, 2023), <https://helpx.adobe.com/photoshop/using/generative-fill.html>.

not liable for hosting those images.⁶²² And this is also precisely the type of situation that indemnification of users could help address.

The main difficulty with this regime would be policing against systems designed specifically for infringement. Something like the *Grokster* rule, carefully followed, might suffice. The providers of a service that was geared to produce infringing outputs could be held liable. So could the publishers or deployers of a model that had been trained or fine-tuned to optimize its effectiveness specifically for infringing uses. So could the curator of a dataset that included only or primarily infringing works, or was intentionally organized to meet the needs of a model known to be intentionally trained for infringement. At every stage, a party would be held responsible only for its own actions specifically directed towards increasing the use of a system for infringement, with no substantial noninfringing purpose.

Notice and Removal

Third, courts could treat generative-AI services as generally legal in themselves, but require them to respond to knowledge of specific infringements under a *Napster*-like rule. One plausible doctrinal route to this regime would be to treat infringing generations as creating direct liability for users and only indirect liability for service providers. Another would use fair use to shield service providers as long as they took reasonable overall precautions, including responding when they had sufficient knowledge of infringement. And a third would be to find liability but craft an injunction that only required services to act against infringement they were aware of.

Regardless of which of these doctrinal routes a court took, there would be an inevitable gravitational force pulling the provider’s duties towards the duties of a service provider under section 512(c) or (d). This is not because Section 512 applies to generative-AI services. It does not.⁶²³ Instead, the Section 512 doctrines may be a convergence point because courts have now had two decades of experience — which means two decades of precedents — with the Section 512 safe harbors. These precedents have come to set expectations — among copyright owners, in the technology industry, in the copyright bar, and in the judiciary — for what legally “responsible” behavior by an online intermediary looks like. A generative-AI service operator that does not appear to be making a good-faith effort to achieve something like this system may strike a court as intending to induce infringement, not making a good-faith effort to comply with an injunction, etc.

⁶²². *E.g.*, *Sinclair v. Ziff Davis, LLC*, 454 F.Supp.3d 342 (S.D.N.Y. 2020).

⁶²³. *See supra* Part II.G.

If courts do end up recreating a notice-and-takedown regime, they would likely settle on familiar elements: a way for copyright owners to give notice of infringement, block infringing generations on notice, block infringing generations on actual knowledge, block infringing generations on red-flag knowledge, avoid having a business model that directly ties income to infringement, and terminate the abilities of repeat infringers to continue making generations. These would probably not be notices directed to specific generations by named users, which would be difficult to detect and track. Instead, they would involve copyright owners identifying copyrighted works and demanding that the generative-AI service operator prevent generations that are substantially similar to those works. Some of those works might be identified based on known outputs that are recognizably similar to suspected inputs. But others might simply involve copyright owners handing over to service operators large catalogs of works to block, much as they currently do with ContentID on YouTube.

This is a very difficult technical problem. It would be much harder for a generative-AI system to implement than it is for a hosting platform to implement Section 512 compliance. The reason is that a notice directed to a hosting provider under Section 512(c) must include “Identification of the material that is claimed to be infringing . . . and information reasonably sufficient to permit the service provider to locate the material.”⁶²⁴ A valid notice is a roadmap; it tells the hosting provider exactly what to take down to comply. That material already exists, and the hosting provider can compare it to the copyrighted work to verify that they are substantially similar. But a notice to a generative-AI system is a notice against future generations, which may be different from each other and resemble the copyrighted work in different ways. Filtering for this kind of much more inexact match is much harder technically.

That said, matching material against a catalog of copyrighted works is a problem that has been very approximately solved by major social networks, which use perceptual hashing to prevent the upload of various kinds of identified content. Generative-AI companies could at least add similar perceptual-hash-driven filtering to the outputs of their models, but clearly this would only solve part of the problem.⁶²⁵ The challenges of implementing removal for models are even harder. A service can add filters on the input and output sides — monitoring prompts and scanning outputs. It can also fine-tune or

⁶²⁴ 17 U.S.C. § 512(c)(3)(A)(i)(i)(i).

⁶²⁵ See generally Lee, Ippolito & Nystrom et al., *supra* note 408 (using hash-driven duplicate detection); Ippolito, Tramèr & Nasr et al., *supra* note 607 (discussing the drawbacks of exact-duplicate detection).

align the model, or provide it with an overall prompt that instructs the model to respond in ways that reduce its propensity to infringe.

But a model by itself does not implement these controls. The model cannot control how it is prompted or what the user does with the output. The model cannot stop anyone from fine-tuning it to remove its guardrails. Further, there is no simple analogue for takedown in generative-AI models. It remains an active and unsolved area of research to figure out how to remove a particular training example's influence from a model's parameters.⁶²⁶ Absent the ability to do so, the safest bet is to retrain the model from scratch. Due to the time and expense required to retrain a model, it will often be infeasible to retrain it simply to remove infringing works, and completely unworkable to retrain on each new notice.

Courts could respond to this difficulty in one of two ways. If they have sympathy for model trainers, they could apply the *Sony* rule, and hold that it is not infringement to distribute a trained model as a set of parameters (as Stability AI's releases have been). The fact that the model is used by others for infringing purposes would be counterbalanced by the substantial non-infringing uses, leading to immunity under *Sony*. This might not always be an attractive business model, because it might be hard for buyers to monetize these models and because of the ease of copying and further redistributing the models, but it could at least exist legally. And truly open-source models would generally be allowed.

But if courts had less sympathy for model trainers, they might hold that the difficulty of complying with removal notices is not an excuse. On this view, the model trainer chose to create a model that could be used for substantial infringement, and to hopelessly commingle infringing and noninfringing material. If so, then it would generally not be legal to distribute a model that was trained on unlicensed works and had infringing outputs, at least once those works they were based on were pointed out. It would be legal to train a model, but the trainer would need to take care that the model was only deployed in a safe environment with sufficient guardrails to prevent infringement. (This is the approach generally taken by OpenAI, for example.)

In this world, open-source models would be extremely risky. As a result, there would likely be a split between two classes of models. Some proprietary models might train on unlicensed works and be deployed only in closed services with carefully designed guardrails. Open-source models would be trained only on public-domain and openly-licensed works, or be trained using very conservative methods to attempt ensure that extremely little copyrighted material was memorized.

626. See, e.g., Meng, Bau, Andonian & Belinkov, *supra* note 442; Bourtole, Chandrasekaran & Choquette-Choo et al., *supra* note 442.

A notice-and-removal regime also has implications for training datasets. A dataset provider cannot pull back these works for which it receives a notice from others who have already used those works for training. But it can delete the works from the dataset it makes available to others going forward. (For an open-source dataset, or one that has been leaked, this second option may be futile, as others will still have copies of the dataset that they can share.) Compared with a model, it is much easier to remove a work from a training dataset; one searches for the work and removes it. Indeed, one could use exact hashing rather than perceptual hashing and still get substantial efficacy in removing a large number of identified works from the dataset — or, for datasets compiled from web crawls or other sources, remove works by tracing their provenance through into the part of the dataset they have ended up in. This makes datasets comparatively more attractive as removal targets, both because they are upstream from many models and because it is easier to define and enforce enforceable removal obligations.

Infringing Models

A fourth possibility is that courts would hold that some or all generative-AI services are illegal because the models themselves infringe. This outcome is an existential threat to many model trainers and service providers; it essentially makes their operations *per se* copyright infringement. It is also the outcome being sought by the class-action plaintiffs in high-profile lawsuits against OpenAI, Stability AI, and some of their partners. In this regime, the most important component of copyright law would quickly become licensing. Models could only be trained on data that had been licensed from the copyright owners, and the terms under which those models and their generations could be used would have to be negotiated as part of the licensing agreement. Each model would have a fully licensed training dataset, and the question of infringement would not arise except in cases where there were infringing works in the dataset itself or some other failure of quality control somewhere along the supply chain.

B. Lessons

Having discussed what courts and policymakers could do, we now consider what they should do. In keeping with our bottom line — *the generative-AI supply chain is too complicated to make sweeping rules prematurely* — we offer a few general observations about the overall shape of copyright and generative AI that courts and policymakers should keep in mind as they proceed.

First, *copyright touches every part of the generative-AI supply chain*. Every stage from training data to alignment can make use of copyrighted works. Generative AI raises many other legal issues: Can a generative-AI system commit defamation?⁶²⁷ Can a generative-AI system do legal work,⁶²⁸ and should they be allowed to?⁶²⁹ But these issues mainly have to do with the outputs of a generative-AI system. Only copyright pervades every step of the process; only copyright is present every time anyone anywhere in the supply chain makes a decision. Copyright cannot be ignored.

Second, and relatedly, *copyright concerns cannot be localized* to a single link in the supply chain. We have argued, time and time again, that decisions made by one actor can affect the copyright liability of another, potentially far away actor in the supply chain. Whether an output looks like Snoopy or like a generic beagle depends on what images were collected in a dataset, which model architecture and training algorithms are used, how trained models are fine-tuned and aligned, how models are embedded in deployed services, what the user prompts with, etc. Every single one of these steps could be under the control of a different person.

Third, *design choices matter*. Every actor in the generative-AI supply chain is in a position to make choices that affect their copyright exposure, and others'. These are obvious choices about copyright, like whether to train on unlicensed data (which can affect downstream risks), and how to respond to notices that a system is producing infringing outputs (which can affect upstream risks). But subtler architectural choices matter, too. Different settings on a training algorithm can affect how much the resulting model will memorize specific works. Different deployment environments can affect whether users have enough control over a prompt to steer a system towards infringing outputs. Copyright law will necessarily have to engage with these choices — as will AI policy more generally.

Fourth, *fair use is not a silver bullet*. For a time, it seemed that training and using AI models would often constitute fair use. In such a world, AI development is generally a low-risk activity, at least from a copyright perspective. Yes, training datasets and models and systems may all include large

627. Eugene Volokh, *Large Libel Models? Liability for AI Output*, 3 J. FREE SPEECH L. 489 (2023); Jon Garon, *An AI's Picture Paints a Thousand Lies: Designating Responsibility for Visual Libel*, 3 J. FREE SPEECH L. 425 (2023); Nina Brown, *Bots Behaving Badly: A Products Liability Approach to Chatbot-Generated Defamation*, 3 J. FREE SPEECH L. 389 (2023); Derek Bambauer & Mihai Surdeanu, *Authorbots*, 3 J. FREE SPEECH L. 375 (2023); Peter Henderson, *Tatsunori Hashimoto, and Mark Lemley, Where's the Liability in Harmful AI Speech?*, 3 J. FREE SPEECH L. 589 (2023).

628. Jonathan H. Choi, Kristen E. Hickman, Amy Monahan & Daniel Schwarcz, *ChatGPT Goes to Law School*, 2023 J. LEGAL EDUC. (forthcoming 2023).

629. *Mata v. Avianca*, No. 22-cv-1461 (S.D.N.Y. June 22, 2023).

quantities of copyrighted works — but they will never be shown to users. Generative AI scrambles this assumption. The serious possibility that some generations will infringe means that the fair-use analysis at every previous stage of the supply chain is up for grabs again.

Fifth, *generative AI does not make the ordinary business of copyright law irrelevant*. Courts will still need to make plenty of old-fashioned, retail judgments about individual works — e.g., how much does this image resemble Elsa and Anna in particular, rather than generic tropes of fantasy princesses? To decide these cases, courts will need to avoid getting distracted by the shininess of new technologies and chasing after inappropriately categorical new rules. Similarity is similarity, proof of copying is proof of copying, transformation in content is transformation in content. Courts *must* leave themselves room to continue making these retail judgments on a case-by-case basis, responding to the specific facts before them, just as they always have. Perhaps, in the fullness of time, as society comes to understand what uses generative AI can be put to and with what consequences, it will reconsider the very fundamentals of copyright law. But until that day, we must live with the copyright system we have. And that system cannot function unless courts are able to say that some generative-AI systems and generations infringe, and others do not.

Sixth, *analogies can be misleading*. There are plenty of analogies for generative AI ready to hand. A generative-AI model or system is like a search engine, or like a website, or like a library, or like an author, or like any number of other people and things that copyright has a well-developed framework for dealing with.⁶³⁰ These analogies are useful, but we wish to warn against treating any of them as definitive. As we have seen, generative AI is and can consist of many things. It is also literally a generative technology: it can be put to an amazingly wide variety of uses.⁶³¹ And one of the things about generative technologies is that they cause convergence;⁶³² precisely because they can emulate many other technologies, they blur the boundaries between things that were formerly distinct. Generative AI can be like a search engine, and also like a website, a library, an author, and so on. Prematurely accepting one of these analogies to the exclusion of the others would mean ignoring numerous relevant similarities — precisely the opposite of what good analogical reasoning is supposed to do.

630. See *supra* Part I.A (for why generations are not like collages).

631. JONATHAN ZITTRAIN, *THE FUTURE OF THE INTERNET – AND HOW TO STOP IT* (2008) (developing theory of generative technologies).

632. See generally Tejas N. Narechania, *Convergence and a Case for Broadband Rate Regulation*, 37 BERKELEY TECH. L.J. 339 (2022) (discussing convergence caused by the Internet).

IV. CONCLUSION

Our conclusion is simple. “Does generative AI infringe copyright?” is not a question that has a yes-or-no answer. There is currently no blanket rule that determines which participants in the generative-AI supply chain are copyright infringers. The underlying technologies and systems are too diverse to be treated identically, and copyright law has too many open decision points to provide clear answers.

Copyright is not the only, or the best, or the most important way of confronting the policy challenges that generative AI poses. But copyright is here, and it is asking good questions about how generative-AI systems are created, how they work, how they are used, and how they are updated. These questions deserve good answers, or failing that, the best answers our copyright system is equipped to give.