This essay is an attempt to work systematically through the copyright infringement analysis of the generative AI supply chain. Our goal is not to provide a definitive answer as to whether and when training or using a generative AI is infringing conduct. Rather, we aim to map the surprisingly large number of live copyright issues that generative AI raises, and to identify the key decision points at which the analysis forks in interesting ways.
INTRODUCTION

This essay is an attempt to work systematically through the copyright infringement analysis of the generative AI supply chain. Our goal is not to provide a definitive answer as to whether and when training or using a generative AI is infringing conduct. Rather, we aim to map the surprisingly large number of live copyright issues that generative AI raises, and to identify the key decision points at which the analysis forks in interesting ways.

We assume basic familiarity with generative AI systems and terminology. We will focus primarily on transformer-based large language models (LLMs) and on diffusion-based image models, but most of the analysis should apply to other architectures and other types of generated content.

Part I provides a brief breakdown of the stages in what we call the “generative AI supply chain.” Part II, the main body of the essay, works through the copyright analysis.

I. The Generative AI Supply Chain

It is helpful to break down the process of training and using a generative AI model into six stages.

A. Data Creation

All generative-AI models are trained on some type of data: examples of the type of material that they are to generate. Text-generation systems are trained on text, image-generation systems are trained on images, music-generation systems on music, and so on. All of this data must come from somewhere. The most prominent current generative AI models have been trained on data that already existed and was not created specifically for the purpose of AI training. And for AIs that generate potentially copyrightable (or copyright-infringing) material, the training data itself will often include copyrightable expression. GitHub Copilot is trained on copyrighted code, ChatGPT is trained on textual data scraped from the web, Stable Diffusion is trained on images, and so on.

For the most part, it is the copyright owners of these individual training works who are the potential plaintiffs in any copyright infringement suit against actors at other stages of the supply chain. These are the relevant copyrights.
B. Dataset Curation

Individual pieces of training data are useless by themselves. The AI training process requires vast quantities of data to create cutting-edge models, and those vast quantities of data must be arranged into datasets that have recurring, standard structure. Sometimes this process is carried out by the same entities that train the AI models. More commonly, however, the process is split across different actors. Stable Diffusion, for example, is trained on images from datasets curated by the non-profit organization LAION. It is necessary, therefore, to consider the potential liability of dataset curators separately from the potential liability of model trainers.

Note that dataset curation, as just described, will frequently involve “the collection and assembling of preexisting materials or of data that are selected, coordinated, or arranged in such a way that the resulting work as a whole constitutes an original work of authorship.” As such, training datasets can themselves be copyrighted, such that the copying without permission of the dataset as a whole could constitute infringement, separate and apart from infringement on the underlying works the dataset comprises. In practice, however, it appears that most uses of training datasets are licensed – either through a bilateral negotiation or by means of an open-source license offered to the world by the dataset compiler.

A few training datasets include metadata on the provenance of their examples, but many datasets do not. Provenance makes it easier to answer questions about the data sources a model was trained on, which can be relevant to an infringement analysis. Provenance also bears on the ease with which specific material can be located, and if necessary removed, from a dataset.

1. This is not to say individual training examples are always unimportant. Specific pieces of training data can have an out-sized influence on generative AI outputs, compared with other pieces of training data.
4. Lee, Cooper, Grimmelmann & Ippolito, supra note 2.
5. For example, the complaint in Tremblay v. OpenAI, Inc. alleges that ChatGPT was trained on books from infringing “shadow libraries” like Library Genesis. Complaint at ¶ 34, Tremblay v. OpenAI, Inc., No. 3:23-cv-03223 (N.D. Cal. June 28, 2023). But this claim is based on circumstantial evidence, because the datasets it was trained on have not been made public.
C. Model Training

Next comes the training process. The model creator selects a model architecture, a training algorithm (including specific values for the algorithm’s hyperparameters),\(^6\) a dataset to apply the algorithm to, and a seed value for the random choices made during the training. Then they wait for the training process to run, at the end of which they hopefully have a useful model. (In the case of generative AI, “useful” means that it produces good generations as outputs when supplied with new inputs, for whatever value of “good” the model creator prefers.) Training requires a substantial investment of resources, including time, storage, and compute. The dollar cost can range from six to eight figures, depending on the size of the model, the size of the training dataset, the length of the training process, the efficiency of the software and hardware used, and other choices.

The resulting model does not by itself do anything. A large language model, for example, is a collection of weights describing the strengths of connections between nodes in a multilayer network. This data structure can be used as part of a generation process by embedding it in a suitable wrapper program. Such a program typically computes a function from an input (a “prompt”) by using the weights to compute the activations of nodes in the model’s network in response to the input. At the end of this process, the program reads off the values from an appropriate part of the network and treats them as the output (a “generation”).

We emphasize this distinction because it is crucial to what comes next. The model that results from training can used to create generations directly, but it does not have to be. In particular, other actors besides the initial model creator may deploy the model and make it available to users – and either of them, or someone else, can modify the model before it is deployed.

D. Model Fine-Tuning

The process of modifying an already-trained model is typically referred to as “fine-tuning.” The existing model might be trained further on new data drawn from a specific problem domain of interest – e.g., a general-purpose language model could be trained on scientific papers to improve its ability to summarize scientific content. Or the model might be trained with examples of outputs to emulate or avoid, as in the reinforcement learning from human feedback (RLHF) methods used to steer language models towards beneficial

---

\(^6\) This is a confusing but important term of art. “Parameters” are the the values (or weights) in a model that the training process learns: they are the outputs of training. “Hyperparameters,” on the other hand, are inputs to training: they are values that are not learned, but instead are typically set by humans to guide the training process.
behaviors and away from negative ones. The general goal is to take an existing model and make it better along some dimension of interest.

As the name suggests, most fine-tuning aims to leverage a model’s general strengths while optimizing its specific details. If we think of training as transforming data into a model, fine-tuning transforms a model into another model. The initial models are sometimes called “foundation” models to indicate that they are intended as bases for fine-tuning a variety of models.\(^7\)

Sometimes, the creator of a model also fine-tunes it. OpenAI’s ChatGPT models are fine-tuned versions of its GPT 3.5 and 4.0 models. In other cases, another party does the fine-tuning. When a model's weights are publicly released (as Meta has with its Llama models), others can take the model and independently fine-tune them for particular applications. To use a copyright analogy, a fine-tuned model is a derivative of the model it is fine-tuned from.

It is helpful to make the foundation-/fine-tuned model distinction because particular behaviors of interest may manifest differently in the two, and because the parties may have different knowledge of, control over, and intentions towards those behaviors. A foundation-model creator, for example, may attempt to train the model to avoid generating copyright-infringing material. But if that model is publicly released, someone else may attempt to fine-tune the model to remove these anti-infringement guardrails. A full copyright analysis may require treating them differently, and indeed, may require analyzing their conduct in relation to each other.

At the same time, it is important to note that a model is a “foundation” or “fine-tuned” model only in relation to other models. These terms do not capture inherent technical features of a model; they describe instead different processes by which a model can be created. A fine-tuned model could itself be treated as foundation model and further fine-tuned by another party. There could be many links in the fine-tuning part of a supply chain.

\(E. \text{ Application Deployment}\)

There are many ways to deploy a model as part of a working application or service. One useful axis is the degree of customization. The Ollama command-line app, for example, is a trivial wrapper app around various LLMs based on Llama: it takes a user-supplied prompt as input and returns the resulting generation as output. But there are other apps that transform user input in complicated ways. The Sydney interactive AI mode in Bing search, for example, combines the user’s query with extensive instructions to create a

\(^7\) Put differently, fine-tuning continues to train a model, but does so with more specific ends in mind. For this reason, training a foundation model is often call “pre-training,” to distinguish from the training that happens as a part of fine-tuning.
compound prompt.\textsuperscript{8} Ollama adds almost no functionality to what is latent in the models it wraps; Sydney behaves very differently than the GPT model embedded in it.

Another important characteristic distinguishing different generative-AI deployments is the degree to which they are controlled by their developers. This is a familiar spectrum from Internet law, from cloud-hosted services at one end to fully open-source software at the other, with downloadable closed-source apps in between. There are examples in generative AI everywhere along the spectrum. ChatGPT and Midjourney are only available to individual users as hosted services; Draw Things is a user-friendly smartphone app based on Stable Diffusion; Ollama is open-source.

\textit{F. Generation}

Finally we are ready to talk about AI users – the people who actually supply prompts and use the resulting generations. There are three elements to a generation. First, there is the \textit{prompt itself}. Some prompts – “a big dog” – are simple and generic, while others – “a big dog facing left wearing a spacesuit in a bleak lunar landscape with the earth rising in the background as an oil painting in the style of Paul Cezanne high-resolution aesthetic trending on artstation” – are more detailed. Second, there is the \textit{choice of application} (which of course embeds an implicit choice of model). A user typically selects an application with the outputs partially in mind, so that one choice or another can indicate an attitude towards the possibility of infringement. And third, there is the \textit{random seed} that goes into each specific generation. It is typical, for example, for image apps to produce four candidate generations, each starting from a different seed (Dall-E and Midjourney both do this). This element of randomness (and of repetition) bears on both copyrightability and infringement.

As we will see, characterizing the relationship between user and app is one of the critical choice points in a copyright-infringement analysis. There are at least three ways the relationship could be described:

- The user actively drives the generation through choice of prompt, and the application passively responds. On this view, the user is potentially a direct infringer, but the application is like a web host, ISP, or other neutral technological provider.

- The application is active and the user passive. On this view, the user is like a viewer of an infringing broadcast, or the unwitting buyer of a pirated

\textsuperscript{8} This kind of prompt transformation is both a complement to and a substitute for fine-tuning. They are both techniques for steering the behavior of a model.
copy of a book. Primary copyright responsibility lies with the application and possibly with others further upstream in the supply chain.

- The user and the application are active partners in generating infringing outputs. On this view, the user is like a patron who commissions a copy of a painting, and the application is like the artist who executes it. They have a shared goal of creating an infringing work.

We will argue that there is no universally correct characterization. Which of these three is the best fit for a particular act of generation will depend on the application, the prompt, and how the application is marketed.

II. Copyright

The hornbook statement of United States copyright doctrine goes something like this: Certain works of authorship are protected by copyright. A defendant infringes when they violate one of the exclusive rights in section 106 by (a) reproducing, adapting, distributing, performing, or displaying an infringing work that is (b) substantially similar to a copyrighted work.

A. The Exclusive Rights

As relevant here, the reproduction right is triggered when the work is reproduced in “copies,” which are defined as “material objects … in which a work is fixed by any method now known or later developed, and from which the work can be perceived, reproduced, or otherwise communicated, either directly or with the aid of a machine or device.” There is no serious question that the computers on which training datasets, models, and generations are stored are “material objects” satisfying this definition — indeed, the computer is itself just such a “machine or device.” So the assembly of a dataset, the training of a model, or the production of a generation is a “reproduction” within the meaning of copyright law. The analysis could stop here, because every link in the generative-AI supply chain requires reproduction in copies and thus implicates copyright. But for the sake of completeness, a similar analysis applies to the other exclusive rights.

The adaptation right is triggered when the defendant “prepare[s] derivative works based upon the copyrighted work.” Whereas the reproduction right is about new copies of an existing work, the adaptation right is about new works based on an existing work. It is best understood as making clear that copyright in a work extends beyond literal similarity to incorporate changes

of form, genre, and content such as translations, sequels, and film adaptations. A training dataset is probably not a derivative work of any of the works in the dataset; it is more appropriately classified as a compilation “formed by the collection and assembling of preexisting materials.”\(^10\) A model is a good example of material that might or might not be an exact reproduction of the works it was trained on, but is more clearly a derivative work because it is “based on” its training data. And generations are frequently derivative works of works in the training data, although whether and when a generation is a derivative of any particular work depends on similarity, discussed below.

The distribution right applies when the defendant “distribute[s] copies … to the public by sale or other transfer of ownership.”\(^11\) Internet-era caselaw confirms that downloads and peer-to-peer transfers infringe the distribution right, so that the essence of the right is giving a stranger a copy, whether or not the copy previously existed.\(^12\) Thus, distribution potentially takes place at the edges of the generative AI supply chain, rather than within the nodes. Purely internal copying (e.g., of the sort that takes place during the training process) does not infringe the distribution right. Similarly, when the same party controls both ends of an edge (e.g., when a model creator collects its own training data, or when a model owner creates test generations for its own use), no distribution takes place. Instead, the distribution right is implicated when parties interact. In our model, there are three such kinds of edges: when a dataset is made available to model creators, when a trained model is made available for download (rather than for interactive use through a web interface or API), and when a service generates generations for users on demand. (Technically, the distribution right is not triggered by merely making a work available for download, but only when someone actually downloads it,\(^13\) but in most interesting cases involving generative AI, the making available is followed by an actual distribution.)

The display and performance rights\(^14\) add relatively little to this picture, because both of them typically involve human perception by someone other than the current owner of a copy. Any case involving a display or performance of a generation to the public will almost always also include a public distribution and a reproduction, so the performance and display rights

\(^{10}\) Id.
\(^{11}\) § 106(3).
\(^{13}\) London-Sire Recs., 542 F. Supp. 2d at 172.
\(^{14}\) A display is static in time; a performance is dynamic. The difference is mostly not significant for the examples we discuss, although there will be some wrinkles in cases involving audio and video.
do not make anyone who shares a generation an infringer unless they were one already. As for training datasets, they are made available in bulk, so the individual works in them are essentially never displayed or performed to strangers. Models as such cannot even be displayed or performed in a way that is comprehensible to humans.

B. Substantial Similarity

Substantial similarity is a qualitative, factual, and frustrating question. Two works are substantially similar when “the ordinary observer, unless he set out to detect the disparities, would be disposed to overlook them, and regard their aesthetic appeal as the same.” A common test is a “holistic, subjective comparison of the works to determine whether they are substantially similar in total concept and feel.” This is not a standard that can be reduced to a simple formula that can easily be applied across different works and genres.

In addition, except in clear cases, substantial similarity is typically a jury question. Juries, unlike judges, are not required to provide reasoned elaboration justifying their verdicts. A typical case in which substantial similarity is genuinely contested, therefore, will provide little guidance for future cases. As a result, it is simply impossible to provide clear, accurate, and actionable predictions of substantial similarity in the mine-run of close cases. The necessary evidence does not exist. For practical purposes, within a broad range, substantial similarity is a nondeterministic black box.

Substantial similarity of training data is so obvious that it requires little discussion. Training datasets contain complete literal copies of millions of copyrighted works. Complete literal copying is a paradigm case where substantial similarity is present as a matter of law.

Substantial similarity of outputs is more complex. At one extreme, a generation may be nearly identical to a work in the training data, as when users have been able to cause models to emit near-exact copies of images in they were trained on. At the other extreme, although a generation is in a meaningful sense based on the works in the training dataset, it may not resemble any of those works in detail. A hypothetical viewer asked to compare the output to each work in the training dataset, one at a time, would say that

it is not substantially similar to work 1, not substantially similar to work 2, and so on through work 89,128,097,032. Under existing copyright doctrine, this should probably end the inquiry: the output is not a “copy” of any of those works, and therefore does not infringe on the copyright in any of them, because infringement is assessed on a work-by-work basis.

In between, there will be difficult cases where similarity requires a more case-by-case and factually intensive analysis. For example, an output might resemble a particular input only in a few recognizable aspects – e.g., a text generation copies a few phrases from an article, or an image generation uses the composition and color palette of a painting but with a different subject, or where a musical generation mashes up multiple songs. In other cases, an output will draw broadly on common elements of a particular artist’s style – an illustrator’s linework, or a photographer’s use of light and shadow – but not be a close copy of any specific work of that artist. All of these cases raise two common doctrinal questions: the quantitative threshold of substantial similarity, and the filtration of unprotectable ideas from the expression in those works.

From our zoomed-out perspective, however, the very factual intensity and complexity of these issues paradoxically makes them legally uninteresting. To quote Learned Hand on the idea-expression dichotomy, “Nobody has ever been able to fix that boundary, and nobody ever can.”

Whether a particular generation is substantially similar or not is ultimately a jury question requiring assessment of audiences’ subjective responses to the works. Generative AI will produce cases requiring this lay assessment, and beyond that it is impossible to anticipate in advance all of the possible variations that will arise. So we assume that some generated outputs will infringe from the viewpoint of lay audiences, but that it will not be possible to perfectly predict which ones will.

Substantial similarity of models, on the other hand, is a legally interesting question. A model, as a collection of weights, is different in kind from the types of copyrightable works it was trained on. Models are not themselves human-intelligible. No viewer would say that the model has the same “total concept and feel” as a painting; no reader would say that it is substantially similar to a blog post; and so on. There is therefore a colorable argument that a model is not a copy of any of the works it was trained on. Although the model could be prompted to generate a substantially similar work, it would be the generation that infringes, and not the model.

This is an interesting argument, and we will return to it below. But there is a persuasive counter-argument that a model is a copy, because the Copyright

Act does not require that copies be directly human-intelligible. After all, a Blu-Ray is not directly intelligible by humans, either, but it counts as a “copy” of the movie on it. Indeed, all digital copies are unintelligible. Instead, they are objects “from which the work can be perceived, reproduced, or otherwise communicated … with the aid of a machine or device.” Thus, even if a model is uninterpretable, it might still be possible to “perceive[]” or “reproduce[]” a copyrighted work embedded in its weights through suitable prompting.

On this view, the necessary condition for a model to count as a substantially similar copy of a work is that the model is capable of generating that work as an output. Note that this is direct infringement, not secondary. The theory is not that the generation is an infringing copy and the model is a tool in causing that infringement in the way that a tape-duplicating machine might be a tool in making infringing cassettes. The theory is that the model itself is an infringing copy, regardless of whether that particular generation is ever made. (Alert readers will note the similarity to the debate over whether the mere act of making it available without a download infringes the distribution right.)

Another perspective is that a model is a compilation of (at least some of) the works that it was trained on. If so, then it might be a copy irrespective of whether the model can generate a specific work when used in the normal way. That is because there might be other ways of inspecting the model that are capable of recovering training data. Obviously, this possibility involves some speculation about technological developments, but it is worth emphasizing that as computer scientists develop techniques that improve the interpretability of models, the copyright treatment of models and generations may well change as a result.

C. Proving Copying

Not all similarity is infringing. Some similarities arise for innocent reasons. The defendant and the plaintiff might both have copied from a common pre-
decessor work, and resemble each other because they both resemble the work they were based on. The similarities might consist entirely of accurate depictions of the same preexisting thing, like Grand Central Station at midday, and resemble each other because Grand Central Station resembles itself. The similarities might be purely coincidental. The plaintiff might even have copied from the defendant!

Copyright law therefore requires that the plaintiff prove that the defendant copied from their work, rather than basing it on some other source or creating it anew, an inquiry known as “copying in fact.” This is a factual question. In some cases, there is direct evidence: e.g., the defendant admits copying or there is video of the defendant using tracing paper to copy a drawing. But in many cases, there are two kinds of indirect evidence: proof that the defendant had access to the plaintiff’s work, and examples of “probative” similarities in the works themselves. Access shows that copying was possible, and similarities can rebut alternative innocent theories.

The copying requirement is straightforward for training datasets, mildly interesting for models, and slightly more interesting for generations. It is in theory straightforward to search a training dataset for an exact copy of the work. Because datasets typically involve compilation of existing works rather than the creation of original works, if a work is in the training dataset at all, it will almost certainly be there because it was copied.\textsuperscript{24} Similarly, if a model is substantially similar to works in the training dataset (see above), it is extremely hard to argue that it was not copied from them.

For outputs, however, when a generative model is uninterpretable, it can be difficult to tell whether an output is similar to a work in the training data because it was copied from that work, or because of coincidence. Because of the extremely wide net that AI companies and organizations cast when assembling training datasets, it will often be easy to show access. Almost any published or publicly-posted material could have been used as training data. The only other useful evidence will therefore be degree of similarity, which means, in effect, both probative similarity and substantial similarity are in play and turn on the same facts – so that the probative similarity inquiry does no additional work.

Vyas, Kakade, and Barak argue that it is possible to define a measure of “near access-freeness” for a model and a copyrighted work such that even if the model was trained on the work, its outputs will be indistinguishable from

\textsuperscript{24} But note that it is a much harder problem to search a large dataset for nonexact copies of a work – such as we might expect if someone else’s derivative of the plaintiff’s work made its way into the training dataset.
a model that was not.\textsuperscript{25} Their model is inspired by copyright’s concept of access, but it is unclear whether copyright law would agree that a near access-free model would actually rebut the evidence that the model was trained on the work in question.

The flip side of these questions is whether a defendant’s showing that a model was not trained on the work in question conclusively disproves copying in fact. From a technical perspective, the argument sounds airtight. The process that led to the allegedly infringing generation is fully documented and entirely independent of the plaintiff’s work – not unlike \textit{Selle v. Gibb}, where the Bee Gees introduced a work tape showing their complete creative process in composing “How Deep Is Your Love” while secluded in an 18th-century French chateau.\textsuperscript{26} The potential fly in the ointment is the evidentiary challenge of actually showing that neither the plaintiff’s work nor any derivatives of it were in the training dataset. Once the possibility of nonexact matches is introduced, the question is substantially harder.

\textbf{D. Direct and Indirect Infringement}

Copyright law has an intricate structure of direct and indirect liability, which is interwoven with its mental elements. Direct liability has no mental requirement: it is “strict liability.” Thus, a person can infringe without intending to — indeed, even without knowing that they are infringing. All that is required that the defendant intentionally made the infringing copy.

Indirect liability comes in three forms. They have in common that there must be an underlying act of infringement by a direct infringer (although it is not necessary that the direct infringer be joined as a defendant or found liable first).

- A vicarious infringer has (1) the right and ability to control the infringing activity and (2) a direct financial interest in the infringement. Vicarious infringement targets parties who have the power to prevent infringement but strong incentives not to. e.g., a swap meet which can expel vendors who sell bootleg music.\textsuperscript{27}


\textsuperscript{26} Selle v. Gibb, 741 F.2d 896, 899 (7th Cir. 1984).

\textsuperscript{27} Fonovisa, Inc. v. Cherry Auction, Inc., 76 F.3d 259, 263 (9th Cir. 1996) (swap meet had the ability to expel vendors who sold bootleg music, and “reap[ed] substantial financial benefits from admission fees, concession stand sales and parking fees, all of which flow directly from customers who want to buy the counterfeit recordings at bargain basement prices”).
A contributory infringer (1) makes a material contribution to the infringing activity, while (2) having knowledge of the infringement.28 Contributory infringement requires parties not to be complicit in infringements they are aware of.

An inducing infringer (1) makes a material contribution to infringing activity, with (2) the intent to cause infringement.29 Contributory infringement – but not vicarious and inducing infringement – is subject to the Sony rule.30 One who distributes a device capable of contributing to infringement – the classic example, from Sony itself is the VCR – is not liable for the resulting infringement, provided that the device is capable of substantial noninfringing uses. Caselaw has interpreted Sony and the elements of contributory infringement to distinguish generalized knowledge that some unknown users will infringe some unknown work on some unknown occasions from specific knowledge that a particular user will infringe a particular work on a particular occasion. The former does not lead to liability; the latter does, provided that the knowledge is obtained before the defendant makes its material contribution. Thus, for example, Napster was not liable for copyright infringements committed by its users unless and until it was on notice of specific infringing songs that it failed to block.31

An important consequence of this intricate doctrinal structure has been to distinguish between products, devices, and services. Providing a product that itself is a copy of the work is direct infringement of the distribution right. Providing a device that can be used to make copies of works is not direct infringement, but can be indirect infringement, subject to the Sony defense. Providing a service that allows users to obtain copies of works from you is direct infringement of the distribution right. Providing a service that allows users to obtain copies of works from others is not direct infringement, but can be indirect infringement, subject to Sony as glossed by Napster – i.e., liability but only on failure to act after notice.32

Finally, there are many cases in which it matters whether a defendant is analyzed as a direct or an indirect infringer. The relevant test here is the “volitional conduct” test,33 which some courts have described in terms of causation: “who made this copy?”34 The direct infringer is the party whose actions toward a specific item of content most proximately caused the infring-
ing activity; anyone else is (potentially) an indirect infringer. Thus, a service that can be used to upload and download infringing content they choose is not a direct infringer, but a service that curates a hand-picked selection of infringing content for users to download is a direct infringer. A copy shop that lets customers operate photocopiers is an indirect infringer; a copy shop that makes the photocopies for them is a direct infringer.

Under this framework, training-set assembly is direct infringement. The curators who select the material for inclusion have made the kind of choices to include certain sources that count as volitional conduct. It does not matter whether they know that specific works are copyrighted; they have chosen to make copies from given sources, and thus they act at their peril under the strict-liability rule. When they make their dataset available for use by others, this is also direct infringement: it is a public distribution.

Similarly, assuming and to the extent that a model is a copy of works in the training dataset, training a model is also direct infringement. Again, the model creators have chosen which datasets to include; they act at their own risk that those datasets include copyrighted material. The same applies to fine-tuning a model: the retrainer takes the risk that they are creating an infringing derivative work. A lack of knowledge as to what works are embedded in the model is no defense. Similarly, the choice to distribute a trained model is sufficiently volitional to count.

The analysis of generation is more complex. Where the same person supplies both the model and the prompt, they are a direct infringer. But when a model owner provides generation as a service to a user who supplies the prompt, the question is which of them is the direct infringer.

By analogy to the copy-shop and UGC service cases, it would seem that the user is the direct infringer and the service provider is the direct infringer. (Imagine, for example, a prompt for “Elsa and Anna from Frozen.”) On this analysis, the service provider is probably not a vicarious infringer, because while they have the right and ability to control their model’s outputs (e.g., by shutting it down or blocking images of Elsa), they do not have enough of a direct financial interest in specifically infringing uses of the service.\(^\text{35}\) They are also typically not an inducing infringer, as they do not intend that the service be used to create infringing cartoons.\(^\text{36}\) As for contributory infringement, the model is a material contribution, but they have only generalized knowledge (some users will make infringing art), not specific knowledge (some

\(^{35}\) This conclusion might be challenged if it were shown that the service’s ability to create infringing generations was a major part of its competitive appeal as compared with other generative-AI services.

\(^{36}\) Again, this conclusion could be challenged if there were sufficient evidence that the service provider did so intend, as was the case in Grokster itself.
users will make art that infringes on *Frozen*). Thus, under *Napster*, they are not liable. A generation service becomes liable, however, when it has specific notice of an infringing work. Once Disney sends a notice to the service over the infringing Elsa output, the service now has the kind of knowledge that triggered liability in *Napster* and must therefore take steps to prevent similar future generations.\(^{37}\) So the bottom line would appear to be that a generation service is liable if and only if it fails to act after receiving notice about specific works.

But this is not the only possible analysis of generation services. Another take on the situation would be that because the model is based on and indeed incorporates copyrighted works as to which the model deployer has the necessary volition, the service is directly liable when it is further used to make infringing outputs. On this view, both the user and the service could be direct infringers. The model is like a very large archive of copyrighted works, so prompting it for a specific generation is like using SciHub to download a specific article.

There is a third possible story. Suppose a user types in “heroic princesses” and the model generates a picture of Elsa and Anna. Here, the user has innocently requested a generation, and it is the model that has narrowed down the enormous space of possible outputs to one that happens to be infringing. There is a colorable argument that the service is a direct infringer, like a bookstore whose shelves are stocked with a mixture of legitimate and pirated editions, but that the user is not. If so, then the user cannot do very much with the output – it is objectively an infringing reproduction and adaptation, so further reproduction and distribution will infringe, regardless of the user’s knowledge – but the generation itself is not directly infringing as to the user. Nor is it indirectly infringing: the user has no financial interest (vicarious) in an infringing output, nor is there knowledge of (contributory) or intent to infringe (inducing).

The two-by-two matrix is not complete. It does not seem likely that a court would treat both service and user as indirect infringers. Doing so would violate the doctrinal requirement that there be a direct infringer for indirect liability to attach, leaving both potentially responsible parties free of liability, and allowing the act of generation to drop out of the copyright system entirely.

\(^{37}\) This too is a contestable conclusion. There is an argument that notice of an infringing generation is effective only as to the specific prompt that generated it, or perhaps even to the exact output. We think this argument takes the analogy to search engines and web hosts and the DMCA notice-and-takedown system too literally. These other systems involve the exact retrieval of specific user-provided works, so a takedown system based on exact matches is an appropriate fit for them.
There is also an argument that models should be treated as devices under *Sony*. This argument comes in two flavors, depending on whether the model is distributed as a set of weights, or made available as a service. The former treats the distribution of a model as distribution of a device for creativity, like a VCR or a computer. This analogy is weak, however, because the VCR and the computer do not come with a library of fragments of copyrighted works embedded within them, and if they generate outputs that are similar to copyrighted works, the information in those outputs came mostly from the model rather than from the prompt.\(^{38}\) The latter version of the argument, which views a generation service itself as a device under *Sony*, is precisely the argument that was rejected in *Napster*.

Finally, a word about content creators. It is theoretically conceivable that the creators of data used to train an AI might be held liable under some secondary liability theory, because their works made possible an AI that infringes on others’ works. But it is highly unlikely, even implausible, that these authors could bear a close enough relationship to the infringing conduct to be held liable solely based on their creation of training data.

### E. Section 512

Section 512 of the Copyright Act, enacted as part of the Digital Millennium Copyright Act, overlays safe harbors for certain online intermediaries on to copyright law. Although these safe harbors have been significant for technology platforms and for Internet law, none of them are likely to apply to generative AI, and for similar reasons. The trainers and operators of a generative AI model will have too much involvement in the model's contents to qualify for these safe harbors.

Section 512(a), which applies to “transient digital network communications,” protects network-level intermediaries like ISPs.\(^ {39}\) It covers only the “transmitting, routing, or providing connections for, material,” and intermediate storage appurtenant thereto, “by or at the direction” of users.\(^ {40}\) This does not describe the way that a model is trained or used. And if there were any doubt, this transmission must occur “through an automatic technical process without selection of the material by the service provider.”\(^ {41}\)


\(^{39}\) 17 U.S.C. § 512(a).

\(^{40}\) 17 U.S.C. § 512(a)(1).

\(^{41}\) Id. § 512(a)(2).
512(a)(2) The comprehensive selection of training data rules out this safe harbor: model trainers choose what model to train on, service providers choose what model to deploy.

Similarly, section 512(b), which covers caching services, is right out. It covers only “intermediate and temporary storage” of “material is made available online by a person other than the service provider.”

Section 512(c), which covers UGC services that store content at the direction of users, is only a slightly nearer miss. It prevents infringement liability “by reason of the storage at the direction of a user of material that resides on a system or network controlled or operated by or for the service provider.” But again, because the service provider here stores material (the model) at its own direction, this is not something that the 512(c) safe harbor covers.

Similarly, section 512(d) prevents liability “by reason of the provider referring or linking users to an online location containing infringing material or infringing activity, by using information location tools, including a directory, index, reference, pointer, or hypertext link.” Again, this is not an apt description of a generative AI model and its generations. The infringing material is coming from the model itself; it is not at some external “online location.”

All of this said, the notice-and-takedown rules under sections 512(c) and 512(d) have been influential enough that they are worth discussing briefly. (They will recur, by way of analogy, later in this paper.) The basic rule is that the safe harbor goes away if the service provider receives a notice about infringing material and fails to disable access to that material. The notice must be specific both about the identity of the copyrighted work being infringed, and about the location where the infringing material is hosted. The point of this regime is to provide the service provider with actionable information that infringement is taking place and how to prevent it. In that sense, it is a codified version of the 

Sony/Napster rule for secondary liability on specific knowledge, together with a mechanism for copyright owners to provide service providers with that knowledge. The model has been so influential that users, platforms, and commentators regularly point to it even in contexts where it does not explicitly apply, e.g. outside the United States, for torts other than copyright infringement, and for platforms that are not themselves eligible for the safe harbors.

There are also a series of other limitations on the safe harbors that create interesting conditions:

42. *Id.* § 512(b).
43. *Id.* § 512(c).
44. *Id.* § 512(d).
45. *Id.* § 512(c)(1)(C).
• The service provider must remove infringing material when it has “actual knowledge” of the infringement.\textsuperscript{46}

• The service provider “is not aware of facts or circumstances from which infringing activity is apparent.”\textsuperscript{47} This language, known as the “red flag” provision, applies in cases where infringement is likely but not absolutely certain.

• The service provider “does not receive a financial benefit directly attributable to the infringing activity, in a case in which the service provider has the right and ability to control such activity”\textsuperscript{48} — a test that looks like the vicarious infringement test, but which courts have interpreted in ways that make it behave more like the specific-knowledge version of the contributory infringement test.

• The service provider must have and enforce a policy for terminating repeat infringers.\textsuperscript{49}

Again, none of these conditions are directly relevant to generative AI models, because they are conditions on a safe harbor that does not apply to these models in the first place. But as we will see, they exert a kind of conceptual force field that warps other doctrines towards resembling them.

\textbf{F. Fair Use}

We have seen that numerous stages of the generative-AI supply chain involve prima facie copyright infringement. This means that copyright’s all-purpose defense, fair use, plays a major role in making generative AI possible at all.\textsuperscript{50} Other authors have discussed the fair use issues in greater detail, so we will focus on only a few salient points.\textsuperscript{51}

First, without generation, there is a strong argument that both assembling training datasets and training AI models is fair use for most applications. The best explanation of this conclusion is Matthew Sag’s concept of nonexpressive uses—bulk uses of copyrighted works that do not involve the

\begin{itemize}
\item \textsuperscript{46} Id. § 512(c)(1)(A)(i).
\item \textsuperscript{47} Id. § 512(c)(1)(B)(i)(i).
\item \textsuperscript{48} Id. § 512(c)(1)(B).
\item \textsuperscript{49} Id. § 512(i)(1)(A).
\item \textsuperscript{50} 17 U.S.C. § 107.
consumption of expression. Examples include digital stylometry, sentiment analysis, and plagiarism detection. These uses do not involve human experience of expression that lies at the heart of the copyright system, and they do not compete with authors. Training a model for these purposes may implicate other important societal interests, but they are not typically described as copyright interests. The reasoning here is essentially backwards-looking. Because the ultimate use does not implicate copyright at all, the intermediate steps of training, fine-tuning, and deploying a model is fair use, and thus the preparatory step of assembling a dataset is also fair use.

This is essentially the logic behind the Google Books fair use decisions. The courts held that the ultimate uses to which the scanned books were put were either fair uses or non-copyright-implicating: provision of books to print-disabled patrons, short (fair use snippets) for search results, and directing users to relevant books. To these we might add the digital humanities research corpus proposed in the (rejected) settlement agreement: other aspects of the settlement attracted vociferous criticism, particularly its treatment of orphan works, but the research corpus was not a principal focus of copyright owners’ objections.

This chain of reasoning, however, does not necessarily follow if an AI model can be used to generate expressive works. We start with the works themselves. The four-factor fair use analysis of such outputs will be case-specific, but we can identify a few broad themes. First, many outputs will be transformative under the first factor. The remixing, abstraction, and yes, transformation of input works seems in a literal sense to involve the kind of generation that Pierre Leval praised in his article introducing the concept: “the quoted matter is used as raw material, transformed in the creation of new information, new aesthetics, new insights and understandings.” Some AI skeptics might deny that AI-generated material can be expressive without a human author, but as long as the audience for these generations finds “new information, new aesthetics, new insights and understandings” in them, the purpose of transformative fair use will be served.

The fact that some generations will be transformative, however, does not mean that all generations will be transformative. In a case where the model substantially memorizes an input work and produces it as an output – whether or not the user’s prompt requested that work – it is hard to argue that

53. Authors Guild v. Google, Inc., 804 F.3d 202 (2d Cir. 2015); Authors Guild, Inc. v. HathiTrust, 755 F.3d 87 (2d Cir. 2014).
the content of the work has been transformed. Nor is there anything about
the context of AI generation to suggest that there is automatically transfor-
mation in the purpose of the use. Some users, at least, will apply these gen-
erations to the same kinds of uses the training data was put to, be they ed-
ucation, entertainment, persuasion, etc. In short, some generations will be
transformative, some will not.

Factor four also is substantially modified for generative AI. The issue here
is that the outputs of a non-generative AI cannot plausibly compete in the
market for the work, but the outputs of a generative AI can. The question
here is whether the generation is a plausible substitute for the original work.
In many cases it will be. Consider the following scenarios:

• The user cannot obtain a particular work at a price they are willing to pay.
  Instead, they ask the model to generate the work, and the model generates
  a close duplicate. The generation is essentially a pirated edition at a lower
  price; it competes with the original for this user’s business.

• The user cannot obtain a particular work at a price they are willing to pay.
  Instead, they ask the model to generate the work, and the model gener-
  ates a non-exact copy with significant aspects borrowed from the original.
  The generation is also a direct competitor under factor four for this user's
  business.

• The user generates a new work in the general genre or style of works by
  a particular author, adapted for their particular needs, e.g., a portrait of
  themself in the style of a named painter. The generation does not compete
  with the original works, which were unsuitable for the user's needs – a por-
 trait of a random stranger is not a substitute for a portrait of me. Instead, it
  competes with commissioning this specific author to produce new work.

• The user generates a new work in a broad style, e.g., “a cyberpunk lion.”
  This generation does not directly compete with any particular work or any
  particular author, but it competes with the works of authors in general. The
  user might have licensed an existing work or commissioned a new work,
  had generative AI not been available.

There are therefore grounds under which a court could find that factor four
favors the copyright owner because the generation harms the legitimate mar-
et for the work. The strength of these grounds will vary from case to case.

It is not possible to make blanket statements about the remainder of fac-
tor one, or about factors two and three.

• Factor one: Some output uses will be commercial, some will not. (We are
  thinking here of the uses made by the users of the generations; there is a
  colorable argument that the commerciality of generation as a service may
be relevant to the fair use claims as to the generations, but we think our analysis is slightly cleaner and more useful.)

- Factor two: Most of the training data will typically have been “published” within the meaning of copyright law; it would not otherwise be available within the training data at all. (A fraction of works may have been held confidentially and released into training datasets without their authors’ permission; the case for removing such works is comparatively stronger.) Some training data will be primarily informational; some will be primarily expressive.

- Factor three: This is a replay of substantial similarity. Some generations will closely resemble the works they were copied from; others will copy comparatively smaller portions of the works, both qualitatively and quantitatively. Even when a work is transformative under the first factor, courts will still also inquire into whether the generation copies more than necessary for that transformation — a “painting of a car driving in a snowstorm in the style of Frida Kahlo” might copy just her color palette, brushwork, and floral motifs, or it might also put the entire composition of *Self-Portrait with Thorn Necklace and Hummingbird* inside the resulting image.

We conclude that some generations will be fair uses and others will not — a conclusion that forces a reconsideration of whether the underlying models are fair uses. The nonexpressive-use argument no longer goes through, because some uses of these models are expressive — and indeed, some of them are infringing.

A full four-factor analysis has been given elsewhere, so we will belabor the details. Instead, we emphasize a few points. First, the models *qua* models are arguably highly transformative — both because they represent the works internally in new and very different ways, and also because they are capable of generating highly transformative works as outputs. Second, the business models of the model trainer and deployer become relevant here: models that are sold or provided as paid services are clearly commercial, whereas open-source releases are not. Third, at least some models will compete with at least some of the authors represented in their training datasets on at least some generations.

Considered as a straight up-or-down question, then, the fair use status of models is a difficult one. One attractive answer for courts may be to hold that the models themselves are fair uses while holding their creators and deployers liable for infringing generations. If generation leads to direct liability, then generation is an at-your-own risk activity. This would make the operation of generation services infeasible unless the operator were able to filter the outputs for all copyrighted works. Doing so would probably require training
only on vetted, licensed datasets (as Adobe’s generative AI model purportedly has been). In this world, however, it would probably not be infringing to distribute a trained model as a set of weights (as Stability AI’s releases have been), because *Sony*. The fact that this tool (which is fair use in itself) is used by others for infringing purposes would be counterbalanced by the substantial noninfringing uses, leading to immunity under *Sony*. This might not 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.

Another response, however, might be to treat generation as creating only indirect liability for model operators. In this world, operators would be allowed to provide models as services – but they would need to respond to notices of specific infringements under a *Napster*-like rule. 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 model 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 model operators large catalogs of works to block, much as they currently do with ContentID on YouTube. Matching new generations against a catalog of copyrighted works is not a trivial problem, but it is one 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. Alternatively, they could retrain the models without the infringing works in the training dataset. This approach may only work on a prospective basis, however, due to the expense and time required to train a full model. It is often infeasible to retrain a model simply to remove infringing works, and it would be completely unworkable to retrain on each new notice.

In practice, there would likely be a gravitational force pulling the operator’s duties towards the duties of a service provider under section 512(c) or (d): 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. It is not the case that these specific requirements arise out of the tests for indirect infringement – the *Napster* rule is only that a provider must block the infringement of specific works as to which it has been provided notice. Instead, we suspect that the section 512 doctrines will be a convergence point because a
model operator that does not implement them might seem to a court to look like an increasingly unappealing candidate for fair use. The reason for this is that 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 model operator that does not appear to be making a good-faith effort to achieve something like this system may strike a court as being irresponsible enough not to warrant a fair use finding. And that conclusion, this judicial horse sense if you will, may well drive the court’s application of the fair-use factors. To be clear, this is speculation on our part. But it seems as plausible a future as several of the others we sketch.

A third possibility is that courts would hold that some or all generative AI models are not fair uses, that 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 world, 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.

Finally, we come to the fair use analysis of the training datasets. Note that a training dataset can be used for training multiple models (indeed, this is overwhelmingly common today) and that not all of those models will be generative ones. The argument against fair use in training datasets therefore depends on there being generations and models that are not fair uses, and it is possible that a model could be unfair even though the dataset it was trained on is fair.

Here is a four-factor analysis of training data:

- Factor one, transformativeness: The transformativeness, if any, in datasets is of a different kind than models and generations. Datasets are not transformative in content; the works may be reformatted and standardized, but there is no new expression. The work itself has been compiled and arranged with other works, but it is unchanged. On the other hand, there is an argument that a dataset for AI training is a transformative purpose: it is a use of a different sort than the usual expressive uses for the work.
itself. Note that both of these points are inflected by the fact that datasets are inputs into AI training, which is transformative in some ways and non-transformative in others.

- Factor one, commerciality: Many training datasets are made publicly available noncommercially. Some observers have argued that this amounts to a kind of ethical and legal laundering by the commercial companies that then train on those datasets – especially when there is a funding relationship between the two. The factor-one commerciality analysis of the dataset may therefore turn on the activities of parties besides the dataset curator.

- Factor two: The dataset will include mostly published works and will include both expressive and informational works, as discussed above.

- Factor three: The dataset typically copies complete works verbatim. This wholesale copying is justified, if at all, in light of the transformative purpose it serves.

- Factor four: In addition to the various markets discussed above, we must consider the market for licensing works for AI training. This is, unfortunately, one of those markets whose definition is entirely circular: this market exists if courts say that it does, and does not exist if courts say that it does not. This factor can be used to justify either fair use or its absence. In previous AI cases, courts have largely found that such markets do not exist, but that reasoning may have been influenced by the fact that they were considering non-generative AIs. With the advent of generative AIs, this question is open again.

One final twist for training datasets concerns removal requests. As with generations, there is an argument that the fair use case may be influenced by whether the dataset curator responds to requests to remove specific works. Note that it cannot pull back these works 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

---


58. *E.g.*, A.V. *ex rel.* Vanderhye v. iParadigms, LLC, 562 F.3d 630 (4th Cir. 2009).
easier to remove a work from a training dataset; one simply 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.

G. Express Licenses

Permission of the copyright owner is a complete defense to infringement; a copyright owner can license any activity they want, thereby rendering it noninfringing. When copyright owners have expressly consented to model training and generation, there is not much further to say. But such licenses apply to only a small fraction of the works currently being used as training data. Some datasets were curated by gathering material from consenting creators expressly for the purpose of model training. In other cases, the copyright agreements that authors have previously executed are so broad that the current owners have the right to allow model training. (This appears to be the case for some stock photography libraries, for example.) Models trained solely with this express and specific permission are rare and smaller and less performant than the most cutting-edge generative-AI models.

An express but non-specific license might arise when works have been posted online under an open license. Here, anyone is allowed to engage in a use as long as it complies with the terms of that license, even if the user has never directly interacted with the copyright owner to obtain specific permission. We will use Creative Commons licenses as an example, as the terms in the Creative Commons license suite cover a useful range of interesting conditions.

Some materials are provided under a Creative Commons Zero notice, which indicates that there are no copyright interests in the material, or that the copyright owner has dedicated the works to the public domain. Any and all uses of these works are allowed, by anyone, without risk of copyright infringement.

The basic license grant in every other Creative Commons license is the right to “reproduce and Share the Licensed Material, in whole or in part; and produce, reproduce, and Share Adapted Material.”

59. Creative Commons Attribution 4.0 International License § 2(a)(1)(A) (2023), https://creativecommons.org/licenses/by/4.0/legalcode.
the section 106 exclusive rights, and it covers all of the activities involved in compiling training datasets, model training and fine-tuning, generation, and use of the generated material. So unless some other license term restricts this grant, generative AI is licensed.

The attribution term in BY licenses requires that the user retain the creator’s identification, indicate whether the work is modified, and retain the Creative Commons license notice. This requirement can be satisfied in “any reasonable manner based on the medium, means, and context.” A training dataset could provide this information through suitable metadata, but many datasets do not. If liability were a serious concern, and the availability of CC-licensed material sufficiently broad to justify it, it is possible that more datasets would bear these attributions, so that they would be fully allowed under CC-BY licenses.

This, however, is where attribution stops with current opaque generative AI models. These models do not attempt to provide any attribution for the works they were trained on. As such, the analysis could stop here. BY is a common term in all of the six standard Creative Commons license, so under none of them is model training allowed. A model that does not retain attribution information cannot provide that information in its generations, so the generations also fall outside a BY license.

The non-commercial term in NC licenses prohibits uses “primarily intended for or directed towards commercial advantage or monetary compensation.” This definition mostly but not entirely tracks the way in which commerciality is defined in fair use, as discussed above. It seems likely that the sale and licensing of datasets and models, and the provision of generations for money would be considered commercial. So this term would allow entirely open-source supply chains, but prohibit any commercial links in those chains.

The no-derivatives term in ND licenses allows the user to copy and share the work itself, but to “produce and reproduce, but not Share, Adapted Material.” Adapted Material is defined as “material … that is derived from or based upon the Licensed Material and in which the Licensed Material is translated, altered, arranged, transformed, or otherwise modified in a manner requiring permission” from the copyright owner. — in other words, any derivative work under copyright law. An ND license therefore allows

60. Id. § 3(a)(1)(A)(i).
61. Creative Commons Attribution-NonCommercial 4.0 International License § 1(i) (2023), https://creativecommons.org/licenses/by-nc/4.0/.
62. Creative Commons Attribution-NoDerivatives 4.0 International License § 2(a)(1)(B) (2023), https://creativecommons.org/licenses/by-nd/4.0/.
63. Id. § 1(a).
dataset curation (as datasets are compilations, not derivatives). But it flatly
prohibits model training and generations as long as they are shared with any-
one else. So one could train models for research, but not share them, and
could not share any generations from that model. The only way for them to
escape from the ND term is for them not to be substantially similar to the
copyrighted work, and thus escape from copyright law entirely. Either way,
an ND license might as well not exist for models and generations.

The share-alike term in SA licenses does allow for the sharing of deriva-
tive works, but they must be placed under the same Creative Commons li-
cense that the underlying works were licensed under. So a model trained
on BY-SA works would itself need to be shared BY-SA, along with any gener-
ations from it. Unlike with BY, this relicensing is feasible without individual
attribution — a blanket BY-SA license applied to a dataset, a model, or a genera-
tion would suffice. (There might still be a failure of the BY term, but the
SA term would be satisfied.)

A lurking issue for all of these cases is that the person who applied a
license to a work might not have the authority to do — they might themselves
be a copyright infringer. In this case, it is hornbook law that the license is
ineffective, and anyone who relies on it is an infringer. There is no defense
of good-faith reliance on a purported license. Improperly licensed materials
can be removed from a dataset; it will be harder to remove them from a model
trained on reliance on them.

H. Implied Licenses

There are also implied copyright licenses, which arise when a copyright owner’s
conduct gives rise to an inference that they have consented to particular uses.
Caselaw holds that the act of putting material online on the web creates an
implied license for search engines to index it and for archives to maintain
archival copies of it. This presumption probably does not apply to mate-
rial behind a paywall or login form, although sites that use such barriers may
also have express licensing in place for datasets based on their data. There is
also some suggestion that this implied license only applies where the owner
has not used a robots.txt file or exclusion headers to deny permission for
bulk crawling. The most prominent training dataset, the Common Crawl,
respects the robots.txt protocol.

64. Creative Commons Attribution-ShareAlike 4.0 International License § 3(b)(1) (2023),
https://creativecommons.org/licenses/by-sa/4.0/.
66. Id. at 1117.
The relevant question, then, is what the scope of this implied license is. If I put a photograph online, with no further information, it is well-established that this act by itself does not grant permission to third parties to use the photograph in news articles or other publications. The implied license allows them to copy the photograph as part of viewing it on my page, but not to use it in other contexts.

A training dataset seems broadly akin to the kind of archives that courts have held to be covered by the implied license in other cases. It is a little harder to say that model training fits within the implied license. This is a new use, one that did not exist when much of the data in the datasets was put online. There is a useful analogy here to the Google Books case – book scanning did not exist when most of the books in the corpus were published, so it is hard to say that authors and publishers consented to scanning when they published.

It is even harder to say that putting material online constitutes an implied license to use that material in AI generations. It is certainly the case that many copyright owners strenuously object to this practice. And if a court is to say that generation is allowed, fair use (which applies whether or not the copyright owner consents) is a better fit for the facts than implied license (which applies only when the copyright owner consents).

This said, the fact that materials were voluntarily placed online can be relevant to the fair use inquiry. As in Sony, which held that taping over-the-air television programs for time-shifting was a fair use, the choice to publish involves giving users access to a work. Copyright owners did not need to license their works for broadcast; they had other alternatives that did not invite the public to view. One would not draw a similar inference from the choice to show a movie in theaters. So even if there is not an implied license as such for AI training, the fact that there is a broadly shared practice of putting material online where any web user can view it is an argument in favor of finding fair use. So implied license inflects the rest of the legal analysis even if it does not directly apply.

67. This point is most clearly seen in the cases holding that news publishers cannot embed photographs posted to Instagram or other social networks. E.g., Sinclair v. Ziff Davis, LLC, 454 F.Supp.3d 342 (S.D.N.Y. 2020).
68. E.g., Field, 412 F. Supp. 2d 1106 (Google Cache).
69. See generally Authors Guild v. Google, Inc., 804 F.3d 202 (2d Cir. 2015).
70. Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417, 456 (1984) (“Sony demonstrated a significant likelihood that substantial numbers of copyright holders who license their works for broadcast on free television would not object to having their broadcasts time-shifted by private viewers.”) (emphasis added).
III. Conclusion

We do not believe that it is currently possible to predict with certainty whether and when participants in the generative-AI supply chain will be held liable for copyright infringement. Our analysis, however, leads to a few general observation about the overall shape of copyright and generative AI.

First, and most importantly, copyright concerns cannot be localized to a single link in the chain. It is taken for granted by some in the AI community that compiling training datasets and training generative models and is fair use. Some outputs might infringe, and that might be a concern for service providers, but on this view the earlier parts of the supply chain are non-infringing. It also certainly appears that some users take the position that copyright is not their problem: if a model is legal to create, it is legal to use. But neither of these propositions is a logical necessity. On the contrary, the question of whether and to what extent upstream or downstream stages involve infringement plays a crucial role in the fair use, secondary liability, and licensing analyses.

Second, as we have argued, there is a critical characterization question of whether to treat the user, the application deployer, or both as a direct infringer. The fact that so much hinges on this question is itself an indictment of the formal and formalistic structure of infringement doctrine. A more rational system, perhaps, would treat alleged direct infringers and alleged secondary infringers more similarly, and would focus instead on the actions and mental states of all parties. But for the moment, as generative-AI infringement lawsuits proceed through the courts, these issues are likely to play an important role.

Third, there are plenty of analogies ready to hand. A generative AI model 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. These analogies are useful, but we wish to warn against treating any of them as definitive. Generative AI is a literally generative technology: it can be put to an amazingly wide variety of uses.71 And one of the things about generative technologies is that they cause convergence.72 Precisely because they can emulate many other technologies, they blur the boundaries between things that were formerly distinct. Generative AI is like a search engine, and also like a website, an author, an author,

---

72. 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).
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.

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, and how they are used. These questions deserve good answers, or failing that, the best answers our copyright system is equipped to give.