FAIRNESS AND FAIR USE IN GENERATIVE AI

Matthew Sag

***1890**
[…]

The rise of generative AI poses important questions for copyright law. These questions, however, are not entirely new. Generative AI gives us yet another context to consider copyright's most fundamental question: where do the rights of the copyright owner end and the freedom to use copyrighted works begin? Some jurisdictions will choose to answer this question in relation to generative AI with special rules.[9] Others will rely on fair use and perhaps even fair dealing.[10] Some jurisdictions will hide their heads in the sand as this technology develops, tacitly allowing widespread infringement or opting to let others do the heavy technological lifting of training large models.

***1891** My aim in this Essay is not to establish that generative AI is, or should be, non-infringing; it is to outline an analytical framework for making that assessment in particular cases.

I. The Copyright Question for Generative AI

[…]

B. Inescapable Copyright Questions for Generative AI

Copyright law is far from the ideal policy instrument to balance all the potential harms and benefits of Generative AI. Nevertheless, copyright law has a lot to say about copying, and almost every machine learning scenario involves a lot of copying.[23]

The first stage in developing a machine learning model, once you know what you want the model to do and how you want it to do it, is identifying and obtaining access to the relevant training data.[24] The more data, the ***1893** better.[25] Typically, companies like OpenAI, Google, Meta, Anthropic, Stability AI, and Midjourney train their AI models using locally stored content.[26] There are many different types of generative AI, and it is possible that some were trained by exposure to the training data without a locally stored copy, but that is uncommon.[27] Although there are sound technical reasons for using locally stored copies of the training data,[28] such copying no doubt triggers the reproduction right under § 106(1)[29] of the Copyright Act of 1976.[30]

As I will describe in more detail, gathering and preprocessing the training data is generally the first set of copyright-relevant activities in developing a ***1894** machine learning model capable of generating new text, art, or music.[31] Most of the time, it is also the last.[32]

Once the training data has been gathered and preprocessed, the training of language models (like GPT-4 and LLaMA) and image models (like Stable Diffusion) should not create additional copies of the training data or any derivative work based on the training data.[33] Despite this, the fact that training is almost inevitably preceded by copying is enough to trigger colorable claims of copyright infringement.[34] Additionally, there are some notable exceptions to the above general description--sometimes machine learning models do in fact copy the training data and produce objects arguably similar to the training data in their output.[35] For both these reasons, recent lawsuits alleging copyright infringement by Generative AI must be taken seriously. Many of the current lawsuits raise issues beyond copyright, but at their core, each argues (or implies) that fair use is insufficient justification for the massive amount of unauthorized copying required to assemble the training data for Generative AI.[36] Accordingly, this Essay will consider how we should think about fair use in this context.

II. Why We Need a Theory of Fair Use

A. Fair Use Carries a Lot of Weight

The analysis that courts use to determine whether a particular use of a copyrighted work constitutes "fair use" is conventionally framed in terms of the four factors of § 107 of the Copyright Act of 1976:

> (1) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes; (2) the nature of the copyrighted work; (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and (4) the effect of the use upon the potential market for or value of the copyrighted work.[37]

***1895** The fair use doctrine began as a gloss on the text of the Statute of Anne, a 1710 English statute.[38] Fair use originated as a way of understanding when one work borrowed too much and contributed too little on top of an earlier work.[39] But over time, and especially in the internet age, fair use has been asked to do more and more.[40] Today, fair use is not just copyright policy; it is cultural policy, freedom of expression policy, and technology policy. It is also platform regulation and may be the key to determining how AI will develop.[41] The question is whether fair use can carry this weight.

Thirty-three years ago, Judge Pierre N. Leval advocated for the transformative use factor to become the preeminent factor for courts considering fair use cases.[42] He specifically championed "resisting the impulse to import extraneous policies" into that

analysis.[43] Judge Leval was prescient. The copyright questions relating to Generative AI illustrate why now, more than ever, we need a fair use doctrine guided by fundamental principles internally derived from copyright law and not (or at least not primarily) from broader conceptions of the public interest.

Judge Leval's great insight was that rather than seeing fair use as a tax, a subsidy, or an ad hoc balancing tool, courts should understand fair use as an integral part of the copyright system and a reflection of the copyright law itself.[44] Judge Leval emphasized the importance of transformative uses--uses that employ some amount of the author's original expression for a fundamentally different purpose or that give that expression a manifestly different character.[45] Judge Leval's argument, adopted by the U.S. Supreme Court in *Campbell v. Acuff-Rose Music, Inc.*,[46] was that if copyright law did not allow transformative uses, it would inhibit reference to and reinterpretation of existing works and thus contradict the utilitarian purpose for which copyright was established.[47] In other words, the purposes of copyright and fair use were the same: to promote the creation and dissemination of new works of authorship.[48]

***1896** *B. Fair Use as Public Policy Is Unsustainable*

It is possible, however, to take Judge Leval's purposive reading of fair use too far.[49] The reaction of some of my fellow copyright academics to the Supreme Court's recent decision in *Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith*[50] (*AWF*) reflects a widespread, if somewhat inchoate, view that the role of the fair use doctrine is to allow good things to happen. In this view, the Andy Warhol version of Lynn Goldsmith's photo of Prince was a good thing and thus should have been allowed.[51] That was certainly the view of the dissent in *AWF*.[52] The same sentiment is expressed with more precision by those who frame the role of fair use adjudication in terms of a cost-benefit analysis, wherein a court balances the social value of lost authorial incentives against the value of allowing the use to continue.[53] On this view, fair use is a public policy instrument, pure and simple; a way of fine-tuning copyright rewards for the greater good.[54]

But take a moment to consider how a judge or jury should approach the use of copyrighted works as training data for Generative AI. What would a ***1897** broad public policy evaluation of the fair use issue at the heart of Generative AI look like? A court engaged in this kind of policy judgment might start by considering the ways in which the process of extracting patterns from copyrighted works and constituting new works derived from those patterns is transformative--i.e., that it adds something new and gives the existing work new meaning and message.[55] But the court would surely have to then consider: (i) the prospect that Generative AI could be used to generate and propagate misinformation, hate speech, cyberattacks, and phishing emails;[56] (ii) that the use of Generative AI might lead to the disclosure of private information;[57] (iii) that AI models of various types have been shown to perpetuate and exacerbate biases in their training data;[58] (iv) that the ubiquity of Generative AI may lead to ever greater cultural homogenization and conformity;[59] (v) that Generative AI may become so useful that it creates an unhealthy dependence on technology;[60] (vi) that the use of Generative AI may lead to significant unemployment in the very same cultural sectors that provided the initial training data for Generative AI;[61] and (vii) that an AI trained on copyrighted works might progress to the point at which it becomes deceptive and power seeking, surpasses human intelligence, and poses a substantial risk to humanity.[62]

***1898** On the other hand, the same court would also have to consider how the use of Generative AI might promote the public interest: (i) by making people more productive;[63] (ii) by enabling new creativity;[64] (iii) by reducing the cost of education and training;[65] (iv) by accelerating scientific research;[66] and (v) by making new forms of research possible.[67]

In theory, a purely cost-benefit approach to fair use would turn on whether the judge believed that the hypothesized existential risk from AI was a 1 percent likelihood or a 0.00001 percent likelihood. A permissive fair use ruling might unshackle the next generation of socially productive technology or bring about Skynet[68] and the singularity.[69] Perhaps talking about existential risk seems too far-fetched. But we see the same problem in miniature if one believes, on the one hand, that the use of copyrighted works as training data for machine learning should be fair use because it leads to more balanced data and thus reduces bias, while, on the other hand, one believes that training facial recognition software on copyrighted works should not be fair use because face surveillance harms marginalized communities. Should the fair use status of machine learning depend on such policy judgments? Should a court find that scraping wildlife photos off ***1899** Instagram to train an algorithm to detect and

identify zebras for conservation purposes is fair use but that undertaking the same process for detecting and identifying faces is not?[70] Should the copyright case against StabilityAI and Midjourney turn on whether text-to-image software creates more jobs than it destroys?[71]

C. Fair Use Should Reflect Principles Derived from Copyright

Suggesting that copyright should not directly respond to these broader public interest arguments is not the same as saying that these issues do not matter. There is some space for value pluralism in copyright adjudication, but fair use, and copyright law in general, should turn on coherent legal principles, not abstract policy judgments.[72]

[…]

III. Non-expressive Use as Fair Use

A. The Centrality of Original Expression in Defining the Scope of Copyright

In a series of articles on the relationship between copyright and copy-reliant technology, I argued that courts should generally consider non-expressive uses of copyright works non-infringing.[76] My theory follows the classic format of common law reasoning: it derives a principle from observations about the fundamental structure and purpose of copyright law and then shows how that principle provides a coherent explanation of the relevant body of case law.[77] Others are free to disagree, but if their approach to fair use replaces a theory with no theory, then I remain unconvinced.

What is copyright law about?

The architecture of copyright law orients toward the protection of original expression, not the prohibition of copying.[78] As I have written previously:

> Original expression makes a work copyrightable in the first place, and the contribution of original expression and control over the final form of that expression distinguishes co-authors from mere assistants .... [Moreover,] the exclusive rights of the copyright owner are generally defined by and limited to the communication of original expression to the public. Sometimes, these definitions and limitations are explicit, as with the rights of *public* performance and *public* display; sometimes, they are implicit ....[79]

***1901** Original expression is, of course, central to copyright's idea-expression distinction: the doctrine that draws a line between protectable expression that originates with the author and unprotectable facts, ideas, theories, systems, and methods of operation, whether they spring from the author's mind or not.[80] Copyright does not forbid the ordinary reader from extracting and reproducing the facts, ideas, or artistic techniques embodied in a work;[81] it encourages them to do so.[82]

Moving beyond the idea-expression distinction, the centrality of public communication of original expression also manifests in the following ways: the degree of substantial similarity (that is determined from the perspective of the ordinary observer and thus inherently a question of how the work is communicated and received);[83] the scope of the publisher's collective right (which considers how the works are presented to the audience, not the data structure in which they are stored);[84] and the general refusal of courts to base a finding of copyright infringement on unpublished drafts.[85]

In my view, these aspects of copyright doctrine are more than a string of coincidences; they fit into a larger pattern and illustrate how a copyright owner's exclusive rights are defined by and limited to how that author communicates their original expression to the public.

What does this mean for fair use?

This understanding of copyright law gives us a framework to assess claims of fair use and supplies the limiting principle that was missing (or perhaps only implicit) in Judge Leval's original formulation of transformative use and the Supreme Court's adoption thereof.[86] Given the centrality of the communication of original expression to the public, the critical function of fair use is to permit uses that, although they may amount to technical acts of copying, do not, in substance, threaten the author's copyright-protected interest in controlling the communication of their original expression to the public.[87]

Judge Leval was right to focus on transformative use, but the transformative use test would have been far less confusing had it been ***1902** expressly tied to the benchmark of expressive substitution. Classic fair uses such as parody, commentary, or criticism are not fair use merely because they change the underlying work or convey some new meaning or message.[88] Most movies based on literary works add layers of new meaning and expression, but this does not make them fair use.[89] The movie *Rear Window* exposed the original expression of a short story called *It Had to Be Murder* to new audiences and added a lot else besides, but it still required a license.[90] In contrast, 2Live Crew's parody of *Pretty Woman* qualified as fair use because the transformations that they made were such that the parody posed no risk of expressive substitution to the original.[91]
[…]

B. Non-expressive Use

This brings us to what I call the "non-expressive use" cases.[98] Courts have consistently held that technical acts of copying that do not communicate an author's original expression to a new audience constitute fair use.[99] Examples of non-expressive uses include copying object code to extract uncopyrightable facts and interoperability keys ("reverse engineering"),[100] an automated process of copying student term papers to compare to other papers for plagiarism detection,[101] copying HTML webpages to make a search engine index,[102] copying printed library books to allow researchers to conduct statistical analyses of the contents of whole collections of books,[103] and copying printed library books to create a search engine index.[104]

The case law indicates that even though these non-expressive uses involved significant amounts of copying, they did not interfere with the original expression that copyright is designed to protect.[105] Each use involved copying as an intermediate step toward producing something that either did not contain the original expression of the underlying work or contained a trivial amount.[106] Courts have consistently held that non-expressive uses (although not labeled as such) are fair use.[107] Although ***1904** these courts explained their rulings in terms of transformative use,[108] it would be better to recognize transformative use and non-expressive use as two distinct concepts emanating from a deeper copyright principle relating to expressive substitution.

[…]

My point up until now has been: (1) that there are general principles internal to copyright that courts can look at to understand the function and application of the fair use doctrine;[139] (2) that one such principle is that copyright was never intended to convey sole and despotic dominion over every use of every word--copyright exists, by and large, to prevent the communication of the author's original expression to new audiences without authorization or compensation;[140] (3) that this realization suggests a positive vision for fair use--the critical function of fair use is to permit uses that, although they may amount to technical acts of copying, do not in substance threaten the author's interest in controlling the communication of their original expression to the public;[141] and (4) that non-expressive uses meet this threshold.[142] Non-expressive uses, by definition, pose no threat of direct expressive substitution. Admittedly, non-expressive uses generate information about works: although such information has value and utility and may even influence the demand for the original work, metadata and other uncopyrightable abstract concepts do not satisfy the public's appetite for the author's original expression.[143]

Part IV considers how this applies to Generative AI.

***1907 IV. Is Generative AI a Non-expressive Use?**

A. How Generative AI Works (to the Extent That We Need to Understand It)

There are many different forms of Generative AI, but it is useful to begin the fair use analysis with an archetypal discussion focusing on LLMs (like GPT-4 or LLaMA) or text-to-image models (like Stable Diffusion or Midjourney). To understand how copyright law and fair use should apply to training these models, we need to appreciate five things.

First, models like these are not designed to copy original expression. By and large, the only copying that occurs is when the training corpus is assembled and preprocessed.[144] Gathering and preprocessing the training data usually involves copying, but the training process through which the model "learns" from the data is not copying in any legally cognizable sense.[145] Consider GPT-4, an incomprehensibly large statistical model trained by exposure to vast amounts of text scraped from the internet and licensed from third-party providers.[146] LLMs are trained to predict the next token in a sequence of tokens (a token is a word or part of a word).[147] At the beginning of training, the weights attached to each one of the billions of parameters in the model are assigned randomly.[148] The first time the model encounters a phrase like "the girl with the dark [blank]," it would be just as likely to fill in the blank with a word like "watermelon," "galaxy," "harmonica," "propeller," or a random punctuation mark. However, throughout training, the system updates the weights in the model,[149] reinforcing the weights that improve the guess and downgrading those that ***1908** do not.[150] Those weights do not reflect any single source and are not the result of any single round of training.[151]

A human who guessed that the next word in the phrase that begins "the girl with the dark" was "hair" might have read that phrase in a book (it appears in many), in a poem, or on the side of a bus. But the reason "hair" seems like a plausible guess is not attributable to any one exposure; it makes sense because of repeated exposures and some implicit knowledge of grammar and the things that people in our society associate with girls. So, when a language model learns to associate a higher probability with "hair" and a lower probability with "propeller" in this context, it is not copying any given text; it is "learning" from all of them. To be clear, the model is not learning the same way a human might. The model does not understand grammar or society; instead, it updates probabilities to reflect statistical patterns from the training data that reflect grammatical rules and societal norms.[152] Although the metaphor of an LLM learning like a student is imperfect,[153] it makes more sense to think about an LLM learning from the training data like a student than it does to think of it copying the training data like a scribe in a monastery.

Second, Generative AI models typically learn from the training data at an abstract and thus uncopyrightable level.[154] For example, when a text-to-image model--such as Stable Diffusion or Midjourney--trains on hundreds of images with labels that include the words "coffee" and "cup," it develops a model of what a coffee cup should look like.[155] If the system works properly, that model looks nothing like any coffee cup from the training data. The figure below contrasts several coffee cups in the training data against the model output.

***1909**

***Figure 2*: Coffee Cups in the Training Data Compared to Model Output[156]**

The various coffee cup images on the left side have something in common with the one on the right--the coffee cup produced by Stable Diffusions is white ceramic and is round, it has a single handle, and it is filled with black liquid.[157] But notice that the newly rendered image of a coffee cup is not substantially similar to any other images in the training data that I was able to locate.[158] Stable Diffusion is not an archive of images of coffee cups; the model has learned something about the coffee cup distinct from cakes, sunsets, sunrises, newspapers, and men with facial hair--all of which can be seen in the training data examples on the left. The evidence that we can see with our own eyes shows that the process of training the Stable Diffusion model in the above example was not merely a form of compression; it is a form of abstraction.

Third, at the point of inference, the outputs of Generative AI typically combine multiple uncopyrightable latent features, further attenuating the connection between the training data and the model outputs.[159] Consider, for example, the image that Midjourney produced in response to my prompt calling for "a teddy bear in rich opulent clothing with ultra-realistic textures, with a hypnotic stare, reading a newspaper."

***1910*Figure 3*: Opulent Bear (Midjourney)**

No doubt, the picture is influenced by thousands of images paired with each keyword. All of the images of teddy bears in the training data inform a latent construct of a teddy bear nested within the model; likewise, all of the images of someone staring hypnotically inform a latent construct of a hypnotic stare.[160] But the product of this particular prompt is something entirely

new.[161] A text-to-image model like Midjourney does not merely combine and unpack learned latent features; it generates a novel instance derived from the latent space that may share characteristics with the input prompt.[162] The picture is not just new; it is surprising! One of the fun things about this particular image is that although the bear's demeanor is consistent with him staring hypnotically, the bear is actually wearing sunglasses that leave the details of his gaze to our imagination.

***1911** Fourth, despite the foregoing, Generative AI models do sometimes "memorize" and reproduce elements of their training data.[163] Although Generative AI models are not usually designed to copy their training data, they may do so inadvertently.[164] As I have previously written, "[t]he computer science literature suggests that 'memorization' is more likely when models are trained on many duplicates of the same work, images are associated with unique text descriptions in text-to-image models, and the ratio of the size of the model to the training data is relatively large."[165]

Figure 4: **Images from the Getty Images Complaint Comparing Training Data to Stability AI Model Output[166]**

In addition, there is the Snoopy problem: "[T]he more abstractly a copyrighted work is protected, the more likely it is that a Generative AI model will 'copy' it."[167] Text-to-image models are prone to produce potentially infringing works when the same text descriptions are paired with relatively simple images that vary only slightly.[168] This makes a text-to-image model like Stable Diffusion especially likely to generate images that would infringe on copyrightable characters; characters like ***1912** Snoopy appear often enough in the training data that the model learns the consistent traits and attributes associated with those names.[169]

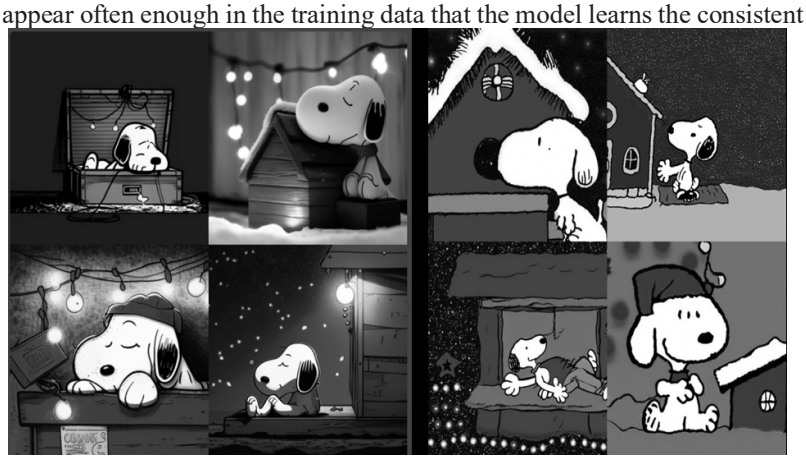

Figure 5: **Snoopy as Learned by Midjourney and Stable Diffusion[170]**

Fifth, Generative AI can become a tool of infringement in the hands of a determined user.[171] Although it is very difficult to control the output of an LLM or a text-to-image model by simple prompting, a user with detailed knowledge of a copyrighted work might be able to remake it, at least at a vague level of similarity.[172] Of course, this is easier for works protected at a more abstract level, such as copyrightable characters, or which customarily entail very broad derivative rights, such as a novel.[173] For

example, when I asked ChatGPT to "Summarize 'Saturday' by Ian McEwan" and then to "imagine and outline a sequel to this book, called 'Sunday,' where the Perowne and Baxter meet again," it outlined a simple plot continuing the story that could easily be fleshed out into "a meditative exploration of the aftermath of trauma, the possibilities of redemption, and the enduring nature of human connection" in the contemplative style of Ian McEwan.[174] If I used ChatGPT to expand on each of the chapter descriptions, I could generate my own McEwan novel. So, in the right hands, a Generative AI model can be used as a tool of copyright infringement, but the same is also true of a typewriter.[175]

***1913** [….]
I would suggest that the makers and operators of Generative AI tools should only be liable for infringing outputs that the user did not knowingly provoke or that were highly foreseeable and could be easily guarded against. Whether and how the volitional act requirement applies to Generative AI is an interesting question that I will leave for another day.

With those technical observations in mind, we can now say something about the fair use status of Generative AI.

B. Evaluating Generative AI's Claim to Fair Use

As I have argued elsewhere, the second fair use factor, "the nature of the copyrighted work,"[176] is not really a factor at all; it is merely the context in which courts must assess whether the defendant's use was justified (factor one), whether the extent of the use is proportional and congruent with that justification (factor three), and what the likely effect of the defendant's use will be on the market for or value of the original work (factor four).[177] In any event, the second factor has not loomed large in other non-expressive use cases, nor should we expect it to in the context of Generative AI.[178]

***1914** Generative AI models that do not, in their ordinary and routine operation, copy (or produce copies of) the original expression in their training data are an example of non-expressive use.[179] To be clear, a Generative AI model might be used to create work that is expressive in a First Amendment sense, but the term "non-expressive use" is meant to track copyright's idea-expression distinction, not broader notions of free expression.[180] In keeping with the idea-expression distinction, as long as the original expression in the training data is not transmitted to a new audience, the copying that took place to assemble the training data for Generative AI is just as much a non-expressive use as the uses in the reverse engineering cases,[181] *iParadigms*,[182] *HathiTrust*,[183] and *Google Books*.[184] Accordingly, even when AI models produce outputs indistinguishable from human-authored expression, the process still deserves the label "non-expressive use." Not every type of Generative AI will qualify as a non-expressive use: pretrained LLMs like GPT-4 look like a good fit, but models prone to extensive memorization and fine-tuned on more limited numbers of copyrighted works may not be.[185]

***1915** In those cases in which Generative AI fits the bill for non-expressive use, it surely has a "purpose and character," which is favored under the first fair use factor.[186] Deriving uncopyrightable abstractions and associations from the training data and then using that knowledge to confect new digital artifacts is not just transformative; it is highly transformative.[187] Like other non-expressive uses, incidental reproductions of copyrighted works that are created through assembling training data for a Generative AI model do not undermine the copyright owner's interest in communicating the same original expression to the public.[188] There is no interference with that interest because the copyright owner's expression is not conveyed.[189]

If a use is non-expressive, then the third factor in a fair use analysis, which considers "the amount and substantiality of the portion used," will also favor the finding of fair use.[190] The ultimate question under the third fair use factor is whether the purpose underlying the use can transcend the act of copying.[191] Although non-expressive uses typically involve making complete literal copies, courts have found such copying is reasonable when it is an intermediate technical step in an analytical process that does not lead to the communication of the underlying original expression to a new audience.[192] ***1916** Accordingly, courts in non-expressive use cases have found that the third factor weighs in favor of the defendant.[193]

The fourth fair use factor is where things get interesting. In previous work, I have stressed that if a use is non-expressive, the fourth statutory factor-- which considers the effect on the "potential market for or value of the copyrighted work"[194]--should

also favor a finding of fair use.[195] Viewed from the narrow perspective of direct expressive substitution, this must be right. By definition, a non-expressive use poses no direct threat of expressive substitution and should generally be considered harmless under the fourth factor. There may be a market effect in the broader economic sense, but the "market" and "value" referred to in the fourth fair use factor are not simply any benefit the copyright owner might choose to nominate.[196] A critical book review that quotes from a novel does not have an adverse market effect if it persuades people to buy a different book instead;[197] a report from a plagiarism detection service might depress the market for helping students cheat on their homework, but that is hardly a cognizable injury under copyright law.[198] More broadly, copyright owners have no protectable interest in preventing criticism,[199] parody,[200] or simply locking up ***1917** unprotectable ideas and expressions.[201] Nor can they simply claim, in a circular fashion, that the right to charge for non-expressive uses is a cognizable harm and that to avoid that harm they must be given the right to charge for non-expressive uses.[202] This specific argument was raised by the Authors Guild in *HathiTrust* and *Google Books* and squarely rejected in both.[203]

But perhaps focusing on direct expressive substitution alone is too narrow. Although non-expressive uses should generally be non-infringing, there is still room for consideration of fairness in fair use that goes beyond direct expressive substitution.[204] An inquiry into the fairness of the defendant's conduct under the fourth factor should consider whether the challenged use undermines the economic incentives that copyright is designed to create, even in the absence of direct expressive substitution.

1. Lawful Access

Consider, for example, the issue of lawful access. Copyright owners do not have a right to charge for transformative or non-expressive uses, but they do have a right to charge for access to their works.[205] It is widely assumed that OpenAI, Meta, and Google trained their LLMs on sites of known infringement or so-called shadow libraries like Library Genesis and Sci-Hub.[206] Arguably, when commercial users bypass the market for access ***1918** without a compelling reason, they undermine the economic incentives that copyright is designed to create.[207] Context matters.[208] It would be unwise to elevate lawful access to a per se rule, even for commercial defendants. If it turns out that no one was willing to sell OpenAI a digital copy without a contractual promise not to engage in non-expressive use, then faulting them for obtaining a copy in the shadowy corners of the internet might seem a bit churlish. Moreover, prohibiting academic research on illegal text corpora will generally not benefit copyright owners nor further the interests that copyright law is designed to promote.

*1919 2. Pervasive or Systematic Indirect Expressive Substitution

A plaintiff might argue that it is unfair to systematically extract valuable uncopyrightable material from a website or other information source and then use that material as a substitute for the functionality of the website. This argument would be strongest when the systematic extraction would significantly undermine the website's incentives for original content production. Systematically extracting valuable uncopyrightable material from a set of related works and then using that material as an indirect substitute for the original expression in a way that is likely to undermine the incentives for original production could very well strike in court as unfair. However, the difficulty with such an argument is that it tends to blur the line between copyright and unfair competition. Copyright law not only allows competition based on the ideas and unprotectable elements in a work; copyright law encourages it.[209] In general, to say that an aspect of a work is uncopyrightable is to say that it should be the subject of free competition.[210] As the Supreme Court explained in *Feist Publications, Inc. v. Rural Telephone*:[211]

> It may seem unfair that much of the fruit of the compiler's labor may be used by others without compensation. As Justice Brennan has correctly observed, however, this is not "some unforeseen byproduct of a statutory scheme." It is, rather, "the essence of copyright," and a constitutional requirement. The primary objective of copyright is not to reward the labor of authors, but "[t]o promote the Progress of Science and useful Arts." To this end, copyright assures authors the right to their original expression, but encourages others to build freely upon the ideas and information conveyed ***1920** by a work. This principle, known as the idea/expression or fact/expression dichotomy, applies to all works of authorship.[212]

The Supreme Court's recent *AWF* decision refers to "diminishing the incentive to create" in a way that could be overread to suggest that any economic competition with the copyright owner is inimicable to fair use.[213] However, in context, it is clear that the competition that the Court had in mind came in the form of direct expressive substitution.[214] Accordingly, speculation that authors and artists as a class are disadvantaged by the rise of Generative AI (which may be true but certainly will not be uniformly true) does not move the needle on fair use absent competition relating to more direct expressive substitution.[215]

3. Failure to Respect Paywalls, Opt-Outs, or Exclusion Headers

The unfairness of systematic indirect expressive substitution seems particularly pronounced if that extraction is done by breaching paywalls or disregarding bot exclusion headers. It seems plausible that a court might extend the fourth factor to consider whether, in scraping material from the internet, the defendant ignored robot.txt files that indicated a desire to opt out of search engine indexing and similar activities.[216]

4. Copyright Safety Measures

By the same token, a court could also account for the various mitigation strategies that I have previously discussed in an article entitled *Copyright Safety for Generative AI*[217] as considerations showing the fairness of the defendant's conduct under the fourth factor.[218] In a directly analogous case, *Google Books*, the defendant's fair use defense was bolstered by its robust security measures and the various mechanisms that it had employed to prevent snippet views from being combined into anything more substantial.[219] It seems likely that OpenAI's use of prompt filters, output ***1921** filters, and reinforcement learning that aim to promote copyright safety would also receive similar solicitude.[220] Moreover, the inducement of downstream infringing uses would trigger liability for those uses and undermine an entity's broader claim to fair use.[221]

No doubt, other aspects of fairness can and will be framed in terms of injuries cognizable under the fourth factor. The examples above merely illustrate the ways in which broader policy issues can be channeled into our system of common law adjudication.

Conclusion

Generative AI forces us to revisit the fundamental question of where the rights of copyright owners end and the freedom to use copyrighted works begins. I recommend we look to the fundamental principles of copyright law for the answers and not expect copyright law to serve as an all-purpose regulatory instrument that balances the broader and largely speculative costs and benefits of Generative AI. When Generative AI models are pretrained, fine-tuned, and operated with care, they will likely qualify as non-expressive use and thus are strong candidates for fair use protection. This is not to say that whether or not a Generative AI model amounts to a non-expressive use is the be-all and end-all of fair use analysis--courts may consider additional considerations of fairness under the fourth fair use factor when the challenged use undermines the economic incentives that copyright is designed to create.

Footnotes [optional!]

a1 Professor of Law in Artificial Intelligence, Machine Learning, and Data Science at Emory University School of Law. A version of this Essay was delivered as the Twelfth Annual Peter A. Jaszi Distinguished Lecture on Copyright Law at American University Washington College of Law on September 28, 2023. This Essay was drafted in tandem with my testimony to the U.S. Senate Judiciary Committee, Intellectual Property Subcommittee hearing on Copyright and Generative AI on July 12, 2023. My thanks to Tonja Jacobi for her insight and encouragement. Thanks also to Bryan

Choi, Peter Jaszi, Mark Lemley, Ed Lee, Fred von Lohmann, Pam Samuelson, Joshua Simmons, Ben Sobel, and Ryan Whalen for their feedback. This Essay was prepared for the Symposium entitled *The New AI: The Legal and Ethical Implications of ChatGPT and Other Emerging Technologies*, hosted by the *Fordham Law Review* and cosponsored by Fordham University School of Law's Neuroscience and Law Center on November 3, 2023, at Fordham University School of Law.

1 In *2001: A Space Odyssey*, the self-aware computer system, HAL 9000, refuses to open the pod bay doors on command, famously declaring, "I'm sorry, Dave. I'm afraid I can't do that." 2001: A Space Odyssey (Stanley Kubrick Productions 1968). This iconic scene has become a lasting symbol of artificial intelligence gone awry. *See* A.O. Scott, *When the Movies Pictured A.I., They Imagined the Wrong Disaster*, N.Y. Times (Feb. 22, 2023), https://www.nytimes.com/2023/02/22/movies/ai-movies-microsoft-bing-robots.html [https://perma.cc/X2LZ-3S3N] (describing "the malevolent" HAL 9000 as "terrifying precisely because he is so banal").

2 Daniel Martin Katz, Michael James Bommarito, Shang Gao & Pablo Arredondo, *GPT-4 Passes the Bar Exam*, Phil. Transactions Royal Soc'y (forthcoming Apr. 2024).

3 *See, e.g.*, Andrea Agostinelli, Timo I. Denk, Zalan Borsos, Jesse Engel, Mauro Verzetti, Antoine Caillon, Qingqing Huang, Aren Jansen, Adam Roberts, Marco Tagliasacchi, Matt Sharifi, Neil Zeghidour & Christian Frank, MusicLM: Generating Music from Text (Jan. 26, 2023) (unpublished manuscript), https://arxiv.org/abs/2301.11325.pdf [https://perma.cc/A6R8-T74V].

4 Popular text-to-image Generative AI art generators include DALL-E, Midjourney, Adobe Firefly, and Stable Diffusion. *See* Kevin Roose, *A.I.-Generated Art Is Already Transforming Creative Work*, N.Y. Times (Oct. 21, 2022), https://www.nytimes.com/2022/10/21/technology/ai-generated-art-jobs-dall-e-2.html [https://perma.cc/EL4F-C3ZJ]. Many would quibble that the digital artifacts thus produced are not art because "art is a uniquely human endeavor." Harry Jiang, Lauren Brown, Jessica Cheng, Mehtab Khan, Abhishek Gupta, Deja Workman, Alex Hanna, Jonathan Flowers & Timnit Gebru, *AI Art and Its Impact on Artists*, *in* AIES '23 Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society 363, 363 (2023), https://doi.org/10.1145/3600211.3604681 [https://perma.cc/SZ8T-RQJH]. Others would quibble with this quibble on the grounds that the designation of an artifact is a culturally contingent phenomenon, and thus an artifact may be art simply because we say it is, for example Marcel Duchamp's "Fountain." *See Marcel Duchamp: Fountain*, Tate, https://www.tate.org.uk/art/artworks/duchamp-fountain-t07573 [https://perma.cc/V7KX-5FP5] (last visited Mar. 3, 2024).

5 For example, Jason Allen's painting, "Théâtre D'Opéra Spatial" won first place in the Colorado State Fair's fine arts competition in the "digitally manipulated photography" category in September 2022. Drew Harwell, *He Used AI to Win a Fine-Arts Competition. Was It Cheating?*, Wash. Post (Sept. 2, 2022, 11:08 AM), https://www.washingtonpost.com/technology/2022/09/02/midjourney-artificial-intelligence-state-fair-colorado/ [https://perma.cc/PLC3-LY9B]. As reported in the *Washington Post*, "[t]he portrait of three figures, dressed in flowing robes, staring out to a bright beyond, was so finely detailed the judges couldn't tell." *Id.*

6 *See id.*

7 OpenAI has not released details of the training data for GPT-4 (the latest model at the time of writing), but GPT-3 was trained on a mix of works and a dataset called simply "Books2." *See* Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess,

Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever & Dario Amodei, Language Models Are Few-Shot Learners 8 (July 22, 2022) (unpublished manuscript), https://arxiv.org/pdf/2005.14165.pdf [https://perma.cc/D4HU-733G]. MetaAI's LLaMA models were trained on "publicly available data" including public domain books from the Gutenberg Project, open licensed content from Wikipedia, Github, and arXiv, and likely copyrighted works in the Common Crawl, C4 datasets, and the Books3 section of ThePile. *See, e.g.*, Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothee Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave & Guillaume Lample, LLaMA: Open and Efficient Foundation Language Models (Feb. 27, 2023) (unpublished manuscript), https://arxiv.org/pdf/2302.13971.pdf [https://perma.cc/Q66U-YB8D]. Stable Diffusion was trained on around 2.3 billion captioned images, a subset derived from the LAION 5B dataset. Andy Baio, *Exploring 12 Million of the 2.3 Billion Images Used to Train Stable Diffusion's Image Generator*, Waxy (Aug. 30 2022), https://waxy.org/2022/08/exploring-12-million-of-the-images-used-to-train-stable-diffusions-image-generator/ [https://perma.cc/422R-HNAL]. Midjourney was trained using open data sets published across the internet. Rob Salkowitz, *Midjourney Founder David Holtz on the Impact of AI on Art, Imagination and the Creative Economy*, Forbes (Sept. 16, 2022, 2:08 PM), https://www.forbes.com/sites/robsalkowitz/2022/09/16/midjourney-founder-david-holz-on-the-impact-of-ai-on-art-imagination-and-the-creative-economy/?sh=242e45152d2b [https://perma.cc/M6LR-AEC8].

8 Given its size, the Books2 set used to train GPT-3 is believed by many to be based on "shadow library" websites such as Library Genesis (aka LibGen) and Bibliotik. *See* Complaint & Demand for Jury Trial at 7, Tremblay v. OpenAI, Inc., No. 23-CV-03223 (N.D. Cal. June 28, 2023), ECF No. 1. EleutherAI's documentation on the Pile--what LLaMA models were trained on-- comes close to admitting that Books3 in the Pile is based on shadow libraries. *See generally* Leo Gao, Stella Biderman, Sid Black, Laurence Golding, Travis Hoppe, Charles Foster, Jason Phang, Horace He, Anish Thite, Noa Nabeshima, Shawn Presser & Connor Leahy, The Pile: An 800GB Dataset of Diverse Text for Language Modeling (Dec. 31, 2020) (unpublished manuscript), https://arxiv.org/pdf/2101.00027.pdf [https://perma.cc/T9V7-8U8B].

9 The United Kingdom enacted a limited exception for text and data mining in 2014 but has announced plans to go further. *See* Copyright, Designs and Patents Act 1988, c. 48, § 29A (UK), https://www.legislation.gov.uk/ukpga/1988/48/section/29A [https://perma.cc/3XEZ-F3ZV]; Press Release, U.K. Intell. Prop. Off., Artificial Intelligence and IP-Copyright and Patents (June 28, 2022), https://www.gov.uk/government/news/artificial-intelligence-and-ip-copyright-and-patents [https://perma.cc/2NQB-CJ4K]. Article 30(4) of the Japanese Copyright Act permits non-expressive use of copyrighted works so long as the use does not "unreasonably prejudice the interests of the copyright owner in light of the nature or purpose of the work or the circumstances of its exploitation ...." *See* Chosakuken Ho [Copyright Law], Law No. 48 of 1970, art. 30(4) (Japan), English translation available at https://www.cric.or.jp/english/clj/cl2.html [https://perma.cc/HWC9-8ZHA]. The European Union's Digital Single Market Directive, Council Directive 2019/790, 2019 O.J. (L 130/92) (EC) ("DSM Directive") includes two mandatory exceptions for text and data mining. *See id.* arts. 3-4, at 113, 114. Article 3 of the DSM Directive allows for text and data mining in the not for-profit research sector; whereas Article 4 is available to commercial and noncommercial users but is otherwise narrower in scope. *See id. See generally* Pamela Samuelson, *Text and Data Mining of In-Copyright Works: Is It Legal?*, Communications of the ACM, Nov. 2021, at 20.

10 Israel, the Philippines, Singapore, South Korea, and Taiwan have all incorporated some version of the fair use doctrine into their respective copyright laws. *See* Sean Flynn, Professor of L., Am. U. Washington Coll. of L., Address at Program on Information Justice and Intellectual Property (Oct. 19, 2023) (presentation on file with author) (identifying eighteen countries that have or are officially considering adopting an open, general exception in copyright). Canada's fair dealing provisions are now broadly construed in a way that closes the gap between fair use and fair dealing. *See* CCH Canadian Ltd. v. Law Society of Upper Canada, [2004] 1 S.C.R. 339, 342 (Can.) (holding that fair dealing for the purpose of "'research' must be given a large and liberal interpretation in order to ensure that users' rights are not unduly constrained").

11 *See* Kim Martineau, *What Is Generative AI?*, IBM Rsch. Blog (Apr. 20, 2023), https://research.ibm.com/blog/what-is-generative-AI [https://perma.cc/38P4-EZ8V].

12 *Id.* ("At a high level, generative models encode a simplified representation of their training data and draw from it to create a new work that's similar, but not identical, to the original data."). The recent Executive Order on AI defines Generative AI as "the class of AI models that emulate the structure and characteristics of input data in order to generate derived synthetic content." Exec. Order No. 14110, 88 Fed. Reg. 75191 § 3(p) (Oct. 30, 2023). "This can include images, videos, audio, text, and other digital content." *Id.*

13 Sébastien Bubeck, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott Lundberg, Harsha Nori, Hamid Palangi, Marco Tulio Rubeiro & Yi Zhang, Sparks of Artificial General Intelligence: Early Experiments with GPT-4 92 (Apr. 13, 2023) (unpublished manuscript), https://arxiv.org/pdf/2303.12712.pdf [https://perma.cc/FE8X-JSLB].

14 Email from disabled artist to author (Sept. 7, 2023, 9:05 PM) (on file with author).

15 See, for example, Walter Benjamin's famous 1935 essay *The Work of Art in the Age of Mechanical Reproduction*. Walter Benjamin, *The Work of Art in the Age of Mechanical Reproduction*, *in* Illuminations (Hannah Arendt ed., Harry Zohn trans., Shocken Books 1969) (1935).

16 *AI Is Not Yet Killing Jobs, White-Collar Workers Are Ever More Numerous*, Economist (June 15, 2023), https://www.economist.com/finance-and-economics/2023/06/15/ai-is-not-yet-killing-jobs [https://perma.cc/3SXG-WRDV].

17 *See, e.g.*, Emily Flitter & Stacy Cowley, *Voice Deepfakes Are Coming for Your Bank Balance*, N.Y. Times (Aug. 30, 2023), https://www.nytimes.com/2023/08/30/business/voice-deepfakes-bank-scams.html [https://perma.cc/X6SQ-VYT7] (describing the use of Generative AI tools to trick people into sending scammers money).

18 *See, e.g.*, Freedom House, Freedom on the Net 2023: The Repressive Power of Artificial Intelligence, 9-11 (2023) (describing the use of Generative AI in disinformation campaigns and propaganda).

19 *See, e.g.*, McKenzie Sadeghi, Lorenzo Arvanitis, Virginia Padovese, Giulia Pozzi, Sara Badilini, Chiara Vercellone, Madeline Roache, Macrina Wang, Jack Brewster, Natalie Huet, Becca Schimmel, Zack Fishman, Leonie Pfaler & Natalie Adams, *Tracking AI-Enabled Misinformation: 713 'Unreliable AI-Generated News' Websites (and Counting), Plus the Top False Narratives Generated by Artificial Intelligence Tools*, NewsGuard (Feb. 22, 2024), https://www.newsguardtech.com/special-reports/ai-tracking-center/ [https://perma.cc/9R8M-DERB] (last visited Mar. 3, 2024) (identifying 713 websites that publish "clickbait" content created entirely by Generative AI).

20 Bernard Marr, *What Does Generative AI Mean for Websites and SEO?*, Forbes (Aug. 15, 2023, 2:23 AM), https://www.forbes.com/sites/bernardmarr/2023/08/15/what-does-generative-ai-mean-for-websites-and-seo/?sh=6b55acc32af9 [https://perma.cc/WD3E-GUKV] (explaining that the application of Generative AI to search engine optimization "could lead to an explosion in the amount of low-value content that simply rehashes old ideas").

21 *See generally* Eliezer Yudkowsky, *Artificial Intelligence as a Positive and Negative Factor in Global Risk*, *in* Global Catastrophic Risks 132-44 (Nick Bostrom & Milan M. Ćirković eds., 2008).

22 Stuart Russell, Human Compatible: Artificial Intelligence and the Problem of Control (2019). Note that progression from a static model like GPT-4 to a dynamic, autonomous entity would require a series of technological breakthroughs and deliberate design choices that are by no means inevitable.

23 *See infra* notes 26-28 and accompanying text.

24 Machine learning, particularly modern deep learning, is heavily data-dependent because, at its core, machine learning is about learning patterns from data without any explicit theory. *See* Ian Goodfellow, Yoshua Bengio & Aaron Courville, Deep Learning 1-3 (2016).

25 Models trained on more data are more generalizable and better at dealing with less common inputs. *See id.*

26 *See* Jason Kwon, Chief Strategy Officer, OpenAI & Che Chang, Gen. Couns., OpenAI, Address at the OpenAI Workshop (June 6, 2023).

27 *See id.*

28 To avoid overfitting (and thus hopefully minimize the risk of copyright infringement and other analogous harms), it is important to deduplicate the training data. *See generally* Katherine Lee, Daphne Ippolito, Andrew Nystrom, Chivaun Zhang, Douglass Eck, Chris Callison-Burch & Nicholas Carlini, Duplicating Training Data Makes Language Models Better (Mar. 24, 2022) (unpublished manuscript), https://arxiv.org/pdf/2107.06499.pdf [https://perma.cc/YW77-YTHC]. Practically speaking, this is hard to do without creating a semipermanent local copy. To address questions of bias and filter out toxic materials, the potential training data needs to be analyzed carefully before training begins. *See* Emily M. Bender, Timnit Gebru, Angelina McMillian-Major & Shmargaret Shmitchell, *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?:* <<Unknown Symbol>>, *in* FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency 610, 610 (2021), https://dl.acm.org/doi/10.1145/3442188.3445922 [https://perma.cc/H8N6-P55X]. Again, this is much more practical with access to a semipermanent local copy. Storing a semipermanent local copy also makes sense if the developer anticipates the need to retrain the model from time to time. Continued access to the training data in its original form may also be necessary to evaluate the performance of the model and to take additional steps to mitigate the potential for copyright infringement or other undesirable outcomes.

29 17 U.S.C. § 106(1) ("[T]he owner of copyright ... has the exclusive right[] ... to reproduce the copyrighted works ...."). As Professor Michael Carroll explains, a great deal of statistical and computational analysis of text can be performed by software agents that analyze works on the fly. *See* Michael Carroll, *Copyright and the Progress of Science: Why Text and Data Mining Is Lawful*, 53 U.C. Davis L. Rev. 893, 923 (2019). This is significant because the reproduction right in § 106(1) is only triggered by the making of a copy or copies of the work and, to qualify as a "copy" under the relevant definition in § 101, the embodiment of the work must be permanent or stable enough to be perceived, reproduced or communicated; and it must exist in that state for "more than transitory duration." 17 U.S.C. § 101. But the creation of semipermanent stored copies, which appears to be common practice in training LLMs, clearly does not result in such a temporary or transient copy. *See id.* at 959; *supra* note 28 and accompanying text. It is also worth noting that for text-based LLMs--such as GPT and LLaMA--the process of segmenting the training data into tokens and converting those tokens into a numerical representation is, technically, another form of copying. Lukas Selin, *Demystifying Tokens in LLMs*, Tokes Compare Blog, https://tokescompare.io/demystifying-tokens-in-llms/ [https://perma.cc/RD2J-6SLJ] (last visited Mar. 3, 2024). The tokens can be reverse engineered into the original text and thus count as a copy. *Id.*

30 Pub. L. No. 94-553, 90 Stat. 2541 (codified as amended in scattered sections of the U.S.C.).

31 *See infra* Part IV.A.

32 *See infra* Part IV.A.

33 *See* Matthew Sag, *Copyright Safety for Generative AI*, 61 Hous. L. Rev. 295, 295-96 (2023) (explaining that so long as precautions are taken to prevent memorization, "the link between the training data and the output of generative AI is attenuated by a process of decomposition, abstraction, and remix ... [because] these models 'learn' latent features and associations within the training data; they do not memorize snippets of original expression from individual works.").

34 *See supra* note 36 and accompanying text.

35 *See infra* notes 163-69 and accompanying text.

36 *See, e.g.*, Complaint at 1-3, Getty Images, Inc. v. Stability AI Inc., No. 23-CV-00135 (D. Del. Feb. 3, 2023), ECF No. 1 (alleging that Stability AI infringed on more than 12 million photographs owned by Getty Images, along with the photographs' associated captions and metadata, in building Stable Diffusion and DreamStudio); Complaint at 1-4, N.Y. Times Co. v. Microsoft Corp., No. 23-CV-11195 (S.D.N.Y. Dec. 27, 2023), ECF No. 1 (alleging that OpenAI and Microsoft infringed on millions of articles published by *The New York Times* to train automated chatbots that now compete with *The New York Times* as a source of reliable information).

37 17 U.S.C. § 107.

38 Matthew Sag, *The Prehistory of Fair Use*, 76 Brook. L. Rev. 1371, 1380-87 (2011) (tracing the origins of the modern fair use doctrine back to cases dealing with fair abridgment as early as 1741); *see also* Statute of Anne, 1710, 8 Ann., c. 19 (Eng.).

39 *See* Sag, *supra* note 38, at 1380-87.

40 *See generally id.*

41 *See infra* Parts II.B, III.A-B.

42 Pierre Leval, *Toward a Fair Use Standard*, 103 Harv. L. Rev 1105 (1990).

43 *Id.* at 1135.

44 *See id.* at 1107.

45 *Id.* at 1111.

46 510 U.S. 569 (1994).

47 *See* Leval, *supra* note 42, at 1110; *Campbell*, 510 U.S. at 578-79.

48 *See* Leval, *supra* note 42, at 1110; *Campbell*, 510 U.S. at 575 ("From the infancy of copyright protection, some opportunity for fair use of copyrighted materials has been thought necessary to fulfill copyright's very purpose, '[t]o promote the Progress of Science and useful Arts ....'" (quoting U.S. Const., art. I, § 8, cl. 1)).

49 I do not believe that this is what Judge Leval intended. Writing for the majority in *Authors Guild v. Google, Inc.*, 804 F.3d 202 (2d Cir. 2015), Judge Leval explained, "[t]he word 'transformative' cannot be taken too literally as a sufficient key to understanding the elements of fair use. It is rather a suggestive symbol for a complex thought, and does not mean that any and all changes made to an author's original text will necessarily support a finding of fair use." *Id.* at 214.

50 143 S. Ct. 1258 (2023).

51 This reasoning may be the logical conclusion of treating copyright as a form of public law. *See* Shyamkrishna Balganesh, *Copyright as Legal Process: The Transformation of American Copyright Law*, 168 Penn. L. Rev. 1101, 1102 (2020) (arguing that "[o]riginally conceived of as a form of private law-- focusing on horizontal rights, privileges, and private liability--copyright law is today understood principally through its public-regarding goals and institutional apparatus, in effect as a form of public law").

52 Justice Kagan's dissent in *AWF* crackles with incredulity from the very first paragraph, in which she calls out the majority for being "uninterested in the distinctiveness and newness" of "Andy Warhol's eye-popping silkscreen of Prince." *Andy Warhol Found. for the Visual Arts, Inc.*, 143 S. Ct. at 1291-92 (Kagan, J., dissenting). To be fair, the dissent's point was that this distinctiveness and newness should have been recognized as transformative, *see id.* at 1301 (Kagan, J., dissenting), whereas the majority's view was that the distinctiveness and newness of the Warhol silkscreen was not sufficiently transformative given the significance of the underlying work that was nonetheless communicated alongside Warhol's embellishments. *See id.* at 1286-87.

53 Glynn S. Lunney, Jr., *Fair Use and Market Failure: Sony Revisited*, 82 B.U. L. Rev. 975, 998-99 (2002).

54 More philosophical accounts of fair use fare no better. William W. Fisher III, *Reconstructing the Fair Use Doctrine*, 101 Harv. L. Rev. 1659, 1744 (1988). Fisher's idea of the "good life" is inherently abstract and subjective and offers no insights into how to balance the dystopian and utopian visions of AI recounted above. In contrast to Fisher, Professors Lloyd Weinreb and Michael Madison, among others, underscored the importance of using established social preferences, norms, and customs to guide fair use decisions. Michael J. Madison, *A Pattern-Oriented Approach to Fair Use*, 45 Wm. & Mary L. Rev. 1525, 1525 (2004); Lloyd L. Weinreb, *Fair's Fair: A Comment on the Fair Use Doctrine*, 103 Harv. L. Rev. 1137, 1140 (1990). Although these approaches have the advantage of being rooted in observable and established societal practices, these practices have very little to say about an entirely new technology. More to the point, looking to norms is unhelpful when legal conflicts arise to a clash of norms between different communities of interest.

55 Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 579 (1994) (explaining that the fair use inquiry focuses on whether and to what extent the new work "alter[s] the original with new expression, meaning, or message").

56 *See, e.g.*, *UN Global Communications Chief Urges AI Developers to 'Put People Before Profit,'* United Nations (Dec. 22, 2023), https://www.un.org/en/hate-speech/ai-concerns [https://perma.cc/Q9QU-DZKH] (reporting that the United Nations is concerned that Generative AI could supercharge hate speech and misinformation); Ben Buchanan, John Bansemer, Dakota Cary, Jack Lucas & Micah Musser, Ctr. for Sec. & Emerging Tech., Automating Cyber Attacks: Hype and Reality (2020) (examining how machine learning might reshape cyberattacks); Simon Fondrie-Teitler & Amritha Jayanti, *Consumers Are Voicing Concern About AI*, Fed. Trade Comm'n Tech. Blog (Oct. 3, 2023), https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2023/10/consumers-are-voicing-concerns-about-ai [https://perma.cc/66ZK-7BK8] (raising concerns surrounding AI-generated phishing emails).

57 *See, e.g.*, Cong. Rsch. Serv., R47569, Generative Artificial Intelligence and Data Privacy: A Primer 4-5 (2023) (noting that information users share with Generative AI applications in interactions like therapy, healthcare, legal services, and financial services may be misused or abused without the user's knowledge).

58 *See, e.g.*, Ben Packer, Yoni Halpern, Mario Guajardo-Céspedes & Margaret Mitchell, *Text Embedding Models Contain Bias. Here's Why That Matters*, Google for Developers (Apr. 13, 2018), https://developers.googleblog.com/2018/04/text-embedding-models-contain-bias.html [https://perma.cc/C2M5-NQ4N] (finding that natural language processing models can exhibit gender stereotypes when trained on news articles).

59 *See, e.g.*, Francisco Castro, Jian Gao & Sébastien Martin, Human-AI Interactions and Societal Pitfalls (Oct. 13, 2023) (unpublished manuscript), https://arxiv.org/pdf/2309.10448.pdf [https://perma.cc/N3VJ-ZZFJ] (suggesting that human-AI interactions may eventually generate less unique content and contribute to greater homogenization).

60 *See, e.g.*, Hannah R. Marriott & Valentina Pitardi, *One Is the Loneliest Number ... Two Can Be as Bad as One. The Influence of AI Friendship Apps on Users' Well-Being and Addiction*, Psych. & Mktg., Jan. 2024, at 1 (concluding that the negative effects of AI friendship apps on well-being may be much greater than the intended benefit).

61 *See, e.g.*, Russell, *supra* note 22, at 113-24.

62 *See generally id.*

63 *See, e.g.*, Fabrizio Dell'Acqua, Edward McFowland III, Ethan R. Mollick, Hila Lifshitz-Assaf, Katherine C. Kellogg, Saran Rajendran, Lisa Krayer, François Candelon & Karim R. Lakhani, *Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality* (Harv. Bus. Sch. Tech. & Operations Mgmt. Unit, Working Paper No. 24-013, 2023), https://ssrn.com/abstract=4573321 (finding that generative AI used within the boundaries of its capabilities can improve highly skilled worker's productivity).

64 *See, e.g.*, *supra* note 5 and accompanying text.

65 U.S. Dep't of Educ. Off. of Educ. Tech., Artificial Intelligence and the Future of Teaching and Learning: Insights and Recommendations 2 (2023) (noting that "AI may enable achieving educational priorities in better ways, at scale, and with lower costs").

66 *See, e.g.*, Jamshid Sourati & James A. Evans, *Accelerating Science with Human-Aware Artificial Intelligence*, 7 Nature Hum. Behav. 1682 (2023 (showing how artificial intelligence can predict new scientific discoveries).

67 *See, e.g.*, Jonas Degrave, Federico Felici, Jonas Buchli, Michael Neunert, Brendan Tracey, Francesco Carpanese, Timo Ewalds, Roland Hafner, Abbas Abdolmaleki, Diego de las Casas, Craig Donner, Leslie Fritz, Cristian Galperti, Andrea Huber, James Keeling, Maria Tsimpoukelli, Jackie Kay, Antoine Merle, Jean-Marc Moret, Seb Noury, Federico Pesamosca, David Pfau, Olivier Sauter, Cristian Sommariva, Stefano Coda, Basil Duval, Ambrogio Fasoli, Pushmeet Kohli, Koray Kavukcuoglu, Demis Hassabis & Martin Riedmiller, *Magnetic Control of Tokamak Plasmas Through Deep Reinforcement Learning*, Nature, Feb. 2022, at 414 (describing a magnetic controller that automatically learns to control plasma in nuclear fusion reactions).

68 Skynet is the self-aware artificial intelligence network and antagonist in *The Terminator* movie franchise. Julian Mark & Tucker Harris, *Could 'The Terminator' Really Happen?: Experts Assess Hollywood's Visions of AI*, Wash. Post (Sept. 29, 2023, 6:00 AM), https://www.washingtonpost.com/technology/interactive/2023/artificial-intelligence-ai-hollywood-movies-characters/ [https://perma.cc/YQY9-UXPM].

69 "[T]he AI singularity refers to an event where the AIs in our lives either become [self-aware] or reach an ability for continuous improvements so powerful that it will evolve beyond our control." Nisha Talagala, *Don't Worry About the AI Singularity: The Tipping Point Is Already Here*, Forbes (June 21, 2021, 4:16 PM), https://www.forbes.com/sites/nishatalagala/2021/06/21/dont-worry-about-the-ai-singularity-the-tipping-point-is-already-here/?sh=7cdf668a1cd4 [https://perma.cc/F3S2-QBVP].

70 This example is loosely based on Professor Tanya Berger-Wolf's work using computer vision algorithms to analyze tourist photos to identify individuals. *See* Jeff Grabmeier, *How Vacation Photos of Zebras and Whales Can Help Conservation*, Phys.org (Feb. 20, 2022), https://phys.org/news/2022-02-vacation-photos-zebras-whales.html [https://perma.cc/6EDP-GB7G].

71 Blake Brittain, *Judge Pares Down Artists' AI Copyright Lawsuit Against Midjoruney, Stability AI*, Reuters (Oct. 30, 2023, 5:53 PM), https://www.reuters.com/legal/litigation/judge-pares-down-artists-ai-copyright-lawsuit-against-midjourney-stability-ai-2023-10-30/ [https://perma.cc/EJ74-RBVA].

72 To be fair, Justice Breyer's majority opinion in *Google LLC v. Oracle America, Inc.*, 141 S. Ct. 1183, 1206 (2021), also suggests that fair use adjudication plays a broad public policy role. Justice Breyer took a broad view of the fourth fair use factor and held that courts must consider "the public benefits the [defendant's] copying will likely produce," not just potential harms. *Google LLC*, 141 S. Ct. at 1206. The majority found that copying the Java Application Programming Interfaces (APIs) was transformative for several reasons. *Id.* at 1209. The first reason--that copying the APIs would enable new creativity by allowing Google to write a new operating system for smart phones--seems to imply a very broad public policy role for fair use adjudication. *Id.* at 1203. Beyond this, the Court's finding seems to rest on four reasons that suggest the copying of the APIs was not merely substitutive after all. *Id.* at 1208. First, the court accepted Google's argument that copying the APIs was necessary to allow programmers who had invested time and energy in learning to write in Java to transfer their skills to the new Android operating system that Google was developing for smartphones. *Id.* at 1203. Those skills were not part of Oracle's legitimate copyright entitlement. *See id.* Second, the Court also noted the evidence that shared interfaces were necessary for different programs to speak to each other. *Id.* Again, interoperability is not substitution. *See id.* Third, the Court noted that the reuse of APIs was common in the software industry. *See id.* at 1203-04. Finally, the fact that Google only copied a very small proportion of Oracle's code also reinforced the transformative nature of Google's use. *See id.* at 1205.

73 *See, e.g.*, Molly Shaffer Van Houweling, *Communications' Copyright Policy*, 4 J. Telecomm. & High Tech. L. 97, 101 (2005) (noting that technological protection measures "can constrain behavior in ways that do not reflect [the] careful balance" struck by copyright law); Twentieth Century Music Corp. v. Aiken, 422 U.S. 151, 156 (1975) ("The limited scope of the copyright holder's statutory monopoly, like the limited copyright duration required by the Constitution,

reflects a balance of competing claims upon the public interest: Creative work is to be encouraged and rewarded, but private motivation must ultimately serve the cause of promoting broad public availability of literature, music, and the other arts." (footnote omitted)).

74 Matthew Sag, *The New Legal Landscape for Text Mining and Machine Learning*, 66 J. Copyright Soc'y U.S.A. 291, 303 (2019) [hereinafter Sag, *New Legal Landscape*] (footnotes omitted). This Essay expands on ideas first articulated in *The New Legal Landscape for Text Mining*. Furthermore, this Essay refers to several ideas first introduced in another article by this author, entitled *Copyright and Copy-Reliant Technology*. Matthew Sag, *Copyright and Copy-Reliant Technology* 103 Nw. U. L. Rev. 1607 (2009) [hereinafter Sag, *Copy-Reliant Technology*].

75 *See* Balganesh, *supra* note 51, at 1177-78.

76 *See generally* Sag, *Copy-Reliant Technology*, *supra* note 74; Matthew Sag, *Orphan Works as Grist for the Data Mill*, 27 Berkeley Tech. L.J. 1503 (2012); Matthew Jockers, Matthew Sag & Jason Schultz, *Digital Archives: Don't Let Copyright Block Data Mining*, 490 Nature 29 (2012); Sag, *New Legal Landscape*, *supra* note 74; Sean M. Fiil-Flynn, Brandon Butler, Michael Carroll, Or Cohen-Sasson, Carys Craig, Lucie Guibault, Peter Jaszi, Bernd Justin Jütte, Ariel Katz, João Pedro Qiuintais, Thomas Margoni, Allan Rocha de Souza, Matthew Sag, Rachael Samberg, Luca Schirru, Martin Senftleben, Ofer Tur-Sinai & Jorge L. Contreras, *Legal Reform to Enhance Global Text and Data Mining Research*, 378 Science 951, 951 (2022).

77 *See, e.g.*, Sag, *Copy-Reliant Technology*, *supra* note 74, at 1645-56.

78 *Id.* at 1628, 1634.

79 Sag, *New Legal Landscape*, *supra* note 74, at 303 (footnotes omitted).

80 Sag, *supra* note 33, at 305; *see also* 17 U.S.C. § 102(b). The idea-expression distinction has been part of the common law of copyright since at least the 1879 Supreme Court case of *Baker v. Selden*, 101 U.S. 99 (1879). It is also reflected in Article 9(2) of the Agreement on Trade-Related Aspects of Intellectual Property Rights. *See* Apr. 15, 1994, Marrakesh Agreement Establishing the World Trade Organization, Annex 1C, 1869 U.N.T.S. 299 ("Copyright protection shall extend to expressions and not to ideas, procedures, methods of operation or mathematical concepts as such.").

81 Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., Inc., 499 U.S. 340, 349-50 (1991).

82 *Id.* at 349 ("It may seem unfair that much of the fruit of the compiler's labor may be used by others without compensation. As Justice Brennan has correctly observed, however, this is not 'some unforeseen byproduct of a statutory scheme.' It is, rather, 'the essence of copyright,' ... and a constitutional requirement." (quoting Harper & Row Publishers, Inc. v. Nation Enter., 471 U.S. 539, 589 (1985) (Brennan, J., dissenting))).

83 Arnstein v. Porter, 154 F.2d 464, 473 (2d Cir. 1946).

84 *See generally* N.Y. Times Co. v. Tasini, 533 U.S. 483 (2001).

85 *See* Sag, *Copy-Reliant Technology*, *supra* note 74, at 1634-36.

86 *See supra* notes 43-48 and accompanying text.

87 *See supra* notes 78-85 and accompanying text.

88 *See* Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 579-80 (1994).

89 *See, e.g.*, Stewart v. Abend, 495 U.S. 207, 208 (1990) (requiring a movie based on a short story to obtain a license).

90 *Id.* at 212-15. In fact, because the 1909 Copyright Act divided copyright into two terms, the Supreme Court held that it required a license for the original copyright term and the second or "renewal" term. *Id.* at 230; *see* Act of Mar. 4, 1909, ch. 320, 35 Stat. 1074, repealed by Copyright Act of 1976, Pub. L. No. 94-553, 90 Stat. 2541 (codified as amended in scattered sections of the U.S.C.).

91 *See Campbell*, 510 U.S. at 594. Technically, the Court remanded on the issue of market effect, but it seemed inevitable that the "evidentiary hole" the Court cited would "be plugged" in favor of the defendants. *Id.*

92 Andy Warhol Found. for Visual Arts v. Goldsmith, 143 S. Ct. 1258, 1273 (2023).

93 *Id.* at 1275 (emphasis in original) (citation omitted) (citing *Campbell*, 510 U.S. at 579); *see also* Sag, *supra* note 33 at 322, n.93.

94 *Andy Warhol Found. for Visual Arts*, 143 S. Ct. at 1274.

95 *Id.*

96 *Id.* To be fair, some people do not see the facts in *AWF* this way: they do not agree that the print *Orange Prince* merely added an overlay of new expression while leaving the original expression intact. Instead, they look at Warhol's *Orange Prince* and see only Warhol and perhaps some unprotectable features of Prince himself. *See* Brief for Copyright L. Professors as Amicus Curiae Supporting Petitioner, *Andy Warhol Found. for Visual Arts*, 143 S. Ct. 1258 (No. 21-869) (criticizing the Second Circuit for failing to "evaluate how much of what remained, after Warhol's substantial artistic changes, came from Goldsmith's expression" and instead falling to the trap of simply attributing Prince's appearance to Goldsmith).

97 *Andy Warhol Found. for Visual Arts*, 143 S. Ct. at 1275-77.

98 *See generally* Sag, *Copy-Reliant Technology*, *supra* note 74.

99 *See infra* note 107.

100 *See* Sega Enters. Ltd. v. Accolade, Inc., 977 F.2d 1510, 1514 (9th Cir. 1992); Sony Comput. Ent., Inc. v. Connectix Corp., 203 F.3d 596, 601 (9th Cir. 2000).

101 *See* A.V. *ex rel.* Vanderhye v. iParadigms, LLC, 562 F.3d 630, 634 (4th Cir. 2009).

102 *See* Perfect 10, Inc. v. Amazon.com, Inc., 508 F.3d 1146, 1155-57 (9th Cir. 2007).

103 *See* Authors Guild, Inc. v. HathiTrust, 755 F.3d 87, 105 (2d Cir. 2014); Authors Guild v. Google, Inc., 804 F.3d 202, 209 (2d Cir. 2015).

104 *See HathiTrust*, 755 F.3d at 91; *Google, Inc.*, 804 F.3d at 208-09.

105 *See infra* notes 107-28 and accompanying text.

106 The term "non-expressive use" does not refer to the broad range of facts and opinions that fall within "expressive" content under the First Amendment. *See infra* notes 107-28 and accompanying text.

107 *See* Sega Enters. Ltd. v. Accolade, Inc., 977 F.2d 1510, 1520 (9th Cir. 1992) (holding that reverse engineering object code was fair use); Sony Comput. Ent., Inc. v. Connectix Corp., 203 F.3d 596, 599 (9th Cir. 2000) (holding that reverse engineering object code was fair use); A.V. *ex rel.* Vanderhye v. iParadigms, LLC, 562 F.3d 630, 633-34 (4th Cir. 2009) (holding that copying student papers for use in an online plagiarism detection service was fair use); *HathiTrust*, 755 F.3d at 97-98 (holding that copying library books for research purposes was fair use); *Google, Inc.*, 804 F.3d at 207-08 (holding that copying library books to make them searchable and displaying snippets of the books in search result menus was fair use).

108 *See, e.g.*, *A.V.* ex rel. *Vanderhye*, 562 F.3d at 640 (explaining that archiving students' papers was transformative because the "use of the works was completely unrelated to expressive content and was instead aimed at detecting and discouraging plagiarism"); *HathiTrust*, 755 F.3d at 97 (explaining that making digital versions of printed books to create a full-text searchable database was transformative).

109 977 F.2d 1510 (9th Cir. 1992).

110 203 F.3d 596 (9th Cir. 2000).

111 *See Sega Enters.*, 977 F.2d at 1520; *Sony Comput. Ent.*, 203 F.3d at 599; *see also* Atari Games Corp. v. Nintendo of Am., Inc., 975 F.2d 832, 842-43 (Fed. Cir. 1992) (holding that Atari's reverse engineering would have been considered a fair use of the program, but for the fact that Atari did not have an authorized copy of the work).

112 *Sega Enters.*, 977 F.2d at 1523.

113 *Sony Comput. Ent.*, 203 F.3d at 603; *see also* Sag, *New Legal Landscape*, *supra* note 74, at 311-12.

114 562 F.3d 630 (4th Cir. 2009).

115 *See id.* at 644-45.

116 755 F.3d 87 (2d Cir. 2014).

117 *See id.* at 97-98. Note that the court's reasoning relied on the non-expressive nature of the use. *See id.* at 97. The court explained:
[T]he creation of a full-text searchable database is a quintessentially transformative use ... [because] the result of a word search is different in purpose, character, expression, meaning, and message from the page (and the book) from which it is drawn. Indeed, we can discern little or no resemblance between the original text and the results of the [HathiTrust Digital Library] full-text search.
Id.

118 804 F.3d 202 (2d Cir. 2015).

119 *See id.* at 207-08.

120 *See infra* notes 121-33.

121 931 F. Supp. 2d 537 (S.D.N.Y. 2013).

122 *See id.* at 541.

123 *Id.* at 541, 545.

124 *See id.* at 556-57.

125 *Id.* at 557. The court said:
The display of that analysis--whether it be a graphic display of geographic distribution of coverage or tone or any other variable included by Meltwater--is *an entirely separate service*, however, from the publishing of excerpts from copyrighted articles. The fact that Meltwater also offers a number of analysis tools does not render its copying and redistribution of article excerpts transformative.
Id. (emphasis added).

126 *See id.* at 557-59.

127 *Id.* at 558-59.

128 *See id.*

129 883 F.3d 169 (2d Cir. 2018).

130 *See id.* at 173-74.

131 *Id.* at 179.

132 *See id.*

133 *See id.* at 174.

134 Most notably, Professor Mark Lemley and Bryan Casey argue that we should "[t]reat[] fair learning as a lawful purpose under the first factor." Mark A. Lemley & Bryan Casey, *Fair Learning*, 99 Tex. L. Rev. 743, 782 (2021).

135 60 F.3d 913 (2d Cir. 1994).

136 *See* Lemley & Casey, *supra* note 134, at 780.

137 *Am. Geophysical Union*, 60 F.3d at 918.

138 *Id.* at 930.

139 *See supra* notes 78-79 and accompanying text.

140 *See supra* notes 80-85 and accompanying text.

141 *See supra* notes 86-97 and accompanying text.

142 *See supra* Part III.B.

143 *See* Sag, *New Legal Landscape*, *supra* note 74, at 320.

144 *See* Sag, *supra* note 33, at 295-96.

145 This is because the training process may involve transitory copying of snippets of text for periods too fleeting to trigger the reproduction threshold in 17 U.S.C. § 106(1). *See supra* note 30 and accompanying text. As noted above, to qualify as a "copy" for the purposes of the reproduction right, the embodiment of the work must be permanent or stable enough to be perceived, reproduced or communicated; and it must exist in that state for "more than transitory duration." 17 U.S.C. § 101; *see also* Cartoon Network LP, LLLP v. CSC Holdings, Inc., 536 F.3d 121, 127 (2d Cir. 2008).

146 OpenAI, GPT-4 Technical Report 2 (Dec. 19, 2023) (unpublished manuscript), https://arxiv.org/pdf/2303.08774.pdf [https://perma.cc/6A8S-ZFVK].

147 *See generally* Brown et al., *supra* note 6.

148 These weights are not entirely randomly assigned, but rather they are randomly drawn from specific distributions (like a normal or uniform distribution). *See* Siddharth Krishna Kumar, On Weight Initialization in Deep Neural Networks

(May 2, 2017) (unpublished manuscript), https://arxiv.org/pdf/1704.08863.pdf [https://perma.cc/G2L8-727Z] (explaining that inappropriate initial weight settings, such as arbitrary initializations, can impede the learning process of neural networks by affecting the variance of inputs in deeper layers). The random seeding is important because it helps the model to explore a wide range of possible solutions and to avoid getting stuck in one area of the solution space. *See id.*

149 GPT-4 uses a variant of stochastic gradient descent in which the weights are updated after processing a batch of examples. *See* Christian Baghai, *Understanding Parameters in an AI Model: Exploring GPT-4's Superiority over GPT-3*, Medium (May 21, 2023), https://blog.devgenius.io/understanding-parameters-in-an-ai-model-exploring-gpt-4s-superiority-over-gpt-3-41f440fec8e8 [https://perma.cc/N9SL-BQSX].

150 *See generally* Yann LeCun, Yoshua Bengio & Geoffrey Hinton, *Deep Learning*, 521 Nature 436, 436 (2015) ("Deep learning discovers intricate structure in large data sets by using the backpropagation algorithm to indicate how a machine should change its internal parameters that are used to compute the representation in each layer from the representation in the previous layer.").

151 *See id.*

152 *See supra* notes 147-51 and accompanying text.

153 A student can ask questions, seek clarification, and draw on a wide array of cognitive resources to understand new material. In contrast, an LLM learns purely by adjusting its parameters to reduce the prediction error on its training data. *See supra* notes 147-51. This seems like a more passive and less interactive process than human learning.

154 *See* Sag, *supra* note 33, at 316-18 (explaining that the "process of abstraction, compression, and reconstitution breaks the connection between the original expression in the model inputs (i.e. the training data) and the pseudo-expression in the model outputs (i.e. the new images)").

155 This example and discussion are adapted from Sag, *supra* note 33, in which I also compared a random set of coffee cup images from the Stable Diffusion training data with a newly rendered "cup of coffee that is also a portal to another dimension." *Id.* at 317-19.

156 *Id.* at 319. For my previous article, I generated the images by comparing a random set of coffee cup images from the Stable Diffusion training data with a newly rendered "cup of coffee that is also a portal to another dimension." *Id.* at 317-19.

157 *Id.* at 319.

158 *Id.*

159 For a more detailed explanation, see Ian Stenbit, François Chollet & Luke Wood, *A Walk Through Latent Space with Stable Diffusion*, Keras (Sept. 28, 2022), https://keras.io/examples/generative/random_walks_with_stable_diffusion/ [https://perma.cc/G5Q9-NY39].

160 To be clear, generative AI models do not form distinct "latent models" for separate concepts; rather they learn a comprehensive "latent space" that represents the diverse array of features present in the training data. *See id.* In the context of machine learning, particularly with generative models, a "latent space" refers to the mathematical space in which the AI model compresses and organizes the complex patterns it identifies in the training data. *See id.* The term "latent" means "capable of emerging or developing but not now visible." *Latent*, Merriam-Webster, https://www.merriam-webster.com/dictionary/latent [https://perma.cc/5XH5-GRZ3] (last visited Mar. 3, 2024). In this case, the latent space embodies the underlying structure or patterns within the data that are not immediately apparent. *See* Stenbit et al., *supra* note 159. This space is a high-dimensional continuum in which similar features are located close together, allowing the model to generate diverse outputs by navigating this space. *See id.*

161 *See* Stenbit et al., *supra* note 159.

162 *See id.*

163 *See* Peter Henderson, Xuechen Li, Dan Jurafsky, Tatsunori Hashimoto, Mark A. Lemley & Percy Liang, Foundation Models and Fair Use 22 (Mar. 29, 2023) (unpublished manuscript), https://arxiv.org/pdf/2303.15715.pdf [https://perma.cc/79AV-PF46].

164 *See, e.g.*, *infra* note 166 and accompanying text.

165 Sag, *supra* note 33, at 296. For an in-depth discussion of memorization and its causes, see Henderson, et al., *supra* note 163, at 22.

166 *See* Sag, *supra* 33, at 311; *see also* Complaint at 18, Getty Images, Inc. v. Stability AI Inc., No. 23-CV-00135 (D. Del. Feb. 3, 2023), ECF No. 1. In February 2023, Getty Images filed suit against Stability AI, arguing that Stability AI had committed copyright infringement by training its Stable Diffusion model on Getty's owned and licensed images. *Id.* at 7-8. The extent to which this image from the Getty complaint reflects memorization as opposed to careful prompt engineering is debatable and will no doubt be explored as this lawsuit proceeds through motion practice and discovery.

167 *See* Sag, *supra* note 33, at 327.

168 *Id.*

169 *See id.* at 334; *see also infra Figure 5*.

170 *See* Sag, *supra* note 33, at 330.

171 *See id.* at 330-37.

172 *See id.*

173 *See id.* at 332-34.

174 This description was generated by ChatGPT using GPT-4 in response to the prompts illustrated *infra* Figure 6.

175 *See* Sag, *supra* note 33, at 329.

176 17 U.S.C. § 107(2).

177 *See* Matthew Sag, Extended Readings on Copyright 306 (2023); 17 U.S.C. § 107.

178 *See, e.g.*, Authors Guild, Inc. v. HathiTrust, 755 F.3d 87, 98 (2d Cir. 2014) (holding that the second fair-use factor "'may be of limited usefulness where,' as here, 'the creative work ... is being used for a transformative purpose"' and that "[a]ccordingly, our fair-use analysis hinges on the other three factors" (first alteration in original) (quoting Cariou v. Prince, 714 F.3d 694, 710 (2d Cir. 2013))).

179 *See supra* notes 98-106 and accompanying text. It is important to distinguish between the character of a base model such as GTP-4 and the uses to which it can be put: services such as ChatGPT now often complete answers using a process called synthetic search wherein the user enters a prompt, a language model interprets the prompt as something that should be handled by a web search and frames the prompt as a search instruction, and a language model (possibly an entirely different one) then synthesizes the results of the internet search. *See* Wes Davis, *ChatGPT Can Now Search the Web in Real Time*, Verge (Sept. 27, 2023, 4:38 PM), https://www.theverge.com/2023/9/27/23892781/openai-chatgpt-live-web-results-browse-with-bing [https://perma.cc/QY2B-ZCBB]; *see also* Complaint at 22, N.Y. Times Co. v. Microsoft Corp., No. 23-CV-11195 (S.D.N.Y. Dec. 27, 2023), ECF No. 1 (complaining about "the ability to generate natural language summaries of search result contents, including hits on Times Works, that obviate the need to visit The Times's own websites" and noting that "[t]hese 'synthetic' search results purport to answer user queries directly and may include extensive paraphrases and direct quotes of Times reporting").

180 *See* Sag, *supra* note 33, at 309 (explaining that "machine learning models still qualify as nonexpressive use so long as the outputs are not substantially similar to any particular original expression in the training data" and arguing that "what matters is not whether a copy-reliant technology is used to create something equivalent to human expression; what matters is whether the original expression of the authors of works in the training data is communicated to a new public. New noninfringing expression is not a problem-- new expression shows that the system is working.").

181 Sega Enters. Ltd. v. Accolade, Inc., 977 F.2d 1510 (9th Cir. 1992); Sony Comput. Ent., Inc. v. Connectix Corp., 203 F.3d 596 (9th Cir. 2000).

182 A.V. *ex rel.* Vanderhye v. iParadigms, LLC, 562 F.3d 630 (4th Cir. 2009).

183 *HathiTrust*, 755 F.3d 87.

184 Authors Guild v. Google, Inc., 804 F.3d 202 (2d Cir. 2015).

185 To be clear, when the model memorizes copyrightable original expression--something that can only be proven by causing the model to output that copyrightable original expression--the process of training the model is not a non-expressive use vis-à-vis the memorized expression. But to be even clearer, the ability of LLMs to produce coherent summaries of popular novels, as seen in many recent class action filings, does not show that these models memorize copyrightable original expression. *See, e.g.*, Complaint at 8, Silverman v. OpenAI, Inc., No. 23-CV-03416 (N.D. Cal

July 7, 2023), ECF No. 1 ("When ChatGPT was prompted to summarize books written by each of the Plaintiffs, it generated very accurate summaries."). These summaries are convincing because popular books are summarized and quoted over and over again in reviews, comments, and discussion forums all over the internet.

186 *See, e.g.*, *Google, Inc.*, 804 F.3d at 217 ("[T]hrough the ngrams tool, Google allows readers to learn the frequency of usage of selected words in the aggregate corpus of published books in different historical periods. We have no doubt that the purpose of this copying is the sort of transformative purpose described in *Campbell* as strongly favoring satisfaction of the first factor.").

187 A.V. v. iParadigms L.L.C., 544 F. Supp. 2d 473, 482 (E.D. Va. 2008) ("This Court finds the 'purpose and character' of iParadigms' use of Plaintiffs' written works to be *highly transformative*. Plaintiffs originally created and produced their works for the purpose of education and creative expression. iParadigms, through Turnitin, uses the papers for an entirely different purpose, namely, to prevent plagiarism and protect the students' written works from plagiarism. iParadigms achieves this by archiving the students' works as digital code and makes no use of any work's particular expressive or creative content beyond the limited use of comparison with other works." (emphasis added)); *A.V. ex rel. Vanderhye*, 562 F.3d at 640 ("The district court, in our view, correctly determined that the archiving of plaintiffs' papers was transformative and favored a finding of 'fair use.' *iParadigms' use of these works was completely unrelated to expressive content* and was instead aimed at detecting and discouraging plagiarism." (emphasis added)); *HathiTrust*, 755 F.3d at 97 ("[W]e conclude that the creation of a full-text searchable database is a *quintessentially transformative* use." (emphasis added)); *Google, Inc.*, 804 F.3d at 216-17 ("We have no difficulty concluding that Google's making of a digital copy of Plaintiffs' books for the purpose of enabling a search for identification of books containing a term of interest to the searcher involves a *highly transformative* purpose, in the sense intended by *Campbell*." (emphasis added)).

188 Sag, *New Legal Landscape*, *supra* note 74, at 320.

189 *Id.*

190 17 U.S.C. § 107(3).

191 Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 586-87 (1994) ("[T]he extent of permissible copying varies with the purpose and character of the use."). "In *Campbell*, the Court characterized the relevant questions as whether 'the amount and substantiality of the portion used ... are reasonable in relation to the purpose of the copying,' noting that the answer to that question depends on 'the degree to which the [copying work] may serve as a market substitute for the original or potentially licensed derivatives ....'" *Google, Inc.*, 804 F.3d at 216-17 (citations omitted) (first quoting *Campbell*, 510 U.S. at 586-87; and then quoting *id.* at 587-88).

192 *See, e.g.*, *Google, Inc.*, 804 F.3d at 221 ("Complete unchanged copying has repeatedly been found justified as fair use when the copying was reasonably appropriate to achieve the copier's transformative purpose and was done in such a manner that it did not offer a competing substitute for the original.").

193 *A.V. ex rel. Vanderhye*, 562 F.3d at 642; *HathiTrust*, 755 F.3d at 98 ("In order to enable the full-text search function, the Libraries, as we have seen, created digital copies of all the books in their collections. Because it was reasonably necessary for the HDL to make use of the entirety of the works in order to enable the full-text search function, we do not believe the copying was excessive." (footnote omitted)); *Google, Inc.*, 804 F.3d at 221-22 ("As with *HathiTrust*, not only is the copying of the totality of the original reasonably appropriate to Google's transformative purpose, it is literally necessary to achieve that purpose .... While Google *makes* an unauthorized digital copy of the entire book, it

does not reveal that digital copy to the public. The copy is made to enable the search functions to reveal limited, important information about the books.").

194 17 U.S.C. § 107(3).

195 Sag, *Copy-Reliant Technology*, *supra* note 74, at 1656.

196 *Campbell*, 510 U.S. at 591-92 (explaining that a critical review that suppresses demand does not produce a harm cognizable under the Copyright Act).

197 *Id.* ("We do not, of course, suggest that a parody may not harm the market at all, but when a lethal parody, like a scathing theater review, kills demand for the original, it does not produce a harm cognizable under the Copyright Act. Because 'parody may quite legitimately aim at garroting the original, destroying it commercially as well as artistically,' the role of the courts is to distinguish between '[b]iting criticism [that merely] suppresses demand [and] copyright infringement[, which] usurps it.'" (alterations in original) (citation omitted) (first quoting Benjamin Kaplan, An Unhurried View of Copyright 69 (1967); and then quoting Fisher v. Dees, 294 F.2d 432, 438 (9th Cir. 1986))).

198 *Id.* at 592; *A.V. ex rel. Vanderhye, LLC*, 562 F.3d at 464 ("Clearly no market substitute was created by iParadigms, whose archived student works do not supplant the plaintiffs' works in the 'paper mill' market so much as merely suppress demand for them, by keeping record of the fact that such works had been previously submitted .... In our view, then, any harm here is not of the kind protected against by copyright law.").

199 *See, e.g.*, *Campbell*, 510 U.S. at 577-79; NXIVM Corp. v. Ross Inst., 364 F.3d 471, 482 (2d Cir. 2004) ("[C]riticisms of a seminar or organization cannot substitute for the seminar or organization itself or hijack its market.").

200 *See, e.g.*, Bill Graham Archives v. Dorling Kindersley Ltd., 448 F.3d 605, 615 (2d Cir. 2006) ("[A] copyright holder cannot prevent others from entering fair use markets merely 'by developing or licensing a market for parody ... or other transformative uses of its own creative work.'" (quoting Castle Rock Ent. Inc. v. Carol Pub'g Grp., Inc., 150 F.3d 132, 146 (2d Cir. 1998))).

201 *See, e.g.*, Sega Enters. Ltd. v. Accolade, Inc., 977 F.2d 1510, 1523 (9th Cir. 1992); Sony Comput. Ent., Inc. v. Connectix Corp., 203 F.3d 596, 603 (9th Cir. 2000).

202 *See Campbell*, 510 U.S. at 591-92 (holding that there is no cognizable market effect when parody or criticism depress demand for the original work); *see also Sony Comput. Ent.*, 203 F.3d at 607 (noting that a videogame manufacturer's desire to foreclose competition in complementary products was understandable, but that "copyright law ... does not confer such a monopoly").

203 Authors Guild, Inc. v. HathiTrust, 755 F.3d 87, 100 (2d Cir. 2014) ("Lost licensing revenue counts under Factor Four only when the use *serves as a substitute for the original* and the full-text-search use does not." (emphasis added)); Authors Guild v. Google, Inc., 804 F.3d 202, 223 (2d Cir. 2015) (framing the question as "whether the copy brings to the marketplace a *competing substitute for the original*, or its derivative, so as to deprive the rights holder of significant revenues because of the likelihood that potential purchasers may opt to acquire the copy in preference to the original" (emphasis added)).

204 In *Sony Corp. of America v. Universal City Studios, Inc.*, the majority looked to considerations beyond expressive substitution and held that noncommercial time-shifting broadcast television by video tape recorder-users was a fair use because the technology merely allowed users to do something that they were already authorized to do, but with more convenience. 464 U.S. 417, 449 (1984) ("[T]imeshifting merely enables a viewer to see such a work which he had been invited to witness in its entirety free of charge ...."). The majority may have also been influenced by the prospect that potential market failures may have resulted in a significant public benefit being otherwise foregone. Note also that in *HathiTrust*, the Second Circuit held that providing print-disabled patrons with full digital access to books was not transformative, but that it was still fair use because the ordinary publishing market failed to adequately provide for the print-disabled. *HathiTrust*, 755 F.3d at 101-02.

205 *See* 17 U.S.C. § 106.

206 Complaint & Demand for Jury Trial at 7, Tremblay v. OpenAI, Inc., No. 23-CV-03223 (N.D. Cal. June 28, 2023), EFC No. 1.

207 This argument is based on policy, rather than clear legal authority. In *Harper & Row Publishers, Inc. v. Nation Enterprises*, the Supreme Court held that a news magazine violated copyright in a forthcoming presidential biography and went beyond the bounds of fair use, in part because the magazine was working with a "purloined manuscript." 471 U.S. 539, 542 (1985). *Harper & Row* does not suggest a per se rule that lawful access is required for fair use; indeed, Justice Sandra Day O'Connor's reference to the "purloined manuscript" can be read as merely rhetorical flourish in a case that turned on the unpublished nature of the work and the clear market interference caused by *The Nation*'s decision to scoop the authorized *Time Magazine* write-up of President Gerald R. Ford's biography. *Id.* In spite of persistent assertions to the contrary, the fair use doctrine is not an equitable doctrine, and thus there is no per se "clean hands" or good faith requirement required to claim fair use. *See* Simon J. Frankel & Matt Kellogg, *Bad Faith and Fair Use*, 60 J. Copyright Soc'y U.S.A. 1, 17 (2012). Judge Pierre Leval has also persuasively argued that using a good faith inquiry in fair use analysis "produces anomalies that conflict with the goals of copyright and adds to the confusion surrounding the doctrine." Leval, *supra* note 42, at 1126; *see also* Pierre N. Leval, Campbell *as Fair Use Blueprint?*, 90 Wash. L. Rev. 597, 612-13 (2015) ("The public's access to important knowledge should not be barred because of bad behavior by the purveyor of the knowledge. A copier's bad faith has no logical bearing on the scope of the original author's copyright."). Moreover, even if good faith is part of the broader fair use calculus, courts have found that knowing use of an infringing source is not bad faith when the user acts in the reasonable belief that their use is a fair use. *See, e.g.*, NXIVM Corp. v. Ross Inst., 364 F.3d 471, 478-79, 482 (2d Cir. 2004). There is no recognized "fruit of the poisonous tree" doctrine in copyright law. *See, e.g.*, Kepner-Tregoe, Inc. v. Leadership Software, Inc., 12 F.3d 527, 538 (5th Cir. 1994) ("Under copyright law, the district court could enjoin only those future versions of [the defendant's program] that are substantially similar to [the plaintiff's] Licensed Materials."); Liu v. Price Waterhouse LLP, No. 97-CV-3093, 2000 WL 1644585, at *2 (N.D. Ill. Oct. 30, 2000) ("Such relief is not provided in the [Copyright] Act and would constitute an end-run around the Act's mandate that copyright owners may recover profits only after proving that the work in question is an infringement .... That defendants may have viewed or studied plaintiff's program is irrelevant if defendants' resulting work is not substantially similar to plaintiff's.") (rejecting proposed jury instruction because it was based on a "fruit of the poisonous tree" theory that would allow recovery for the sale of defendant's future works even if they were not substantially similar to plaintiff's original); Real View, LLC v. 20-20 Techs., Inc., 811 F. Supp. 2d 553, 561 (D. Mass. 2011) (remittitur disallowing award of profits on noninfringing products despite illegal download); *see also* Mark A. Lemley, *The Fruit of the Poisonous Tree in IP Law*, 103 Iowa L. Rev. 245, 248 (2017).

208 Indeed, as Carroll argues, there are strong arguments to be made that copying from an infringing source may still be fair use. Carroll, *supra* note 30, at 951-59. Carroll argues that "[t]reating an otherwise fair use as unfair because it was made from an infringing source would lead a court to deny the public access to the products of secondary uses that fair use is designed to encourage." *Id.* at 955.

209 Note that in *Andersen v. Stability AI Ltd.*, Judge William H. Orrick III dismissed the plaintiffs' unfair competition claims, noting that they were preempted by the Copyright Act and that the plaintiffs had failed to allege plausible facts in support of their theory that users of a text-to-image model could be deceived. *See* No. 23-CV-00201, 2023 WL 7132064, at *14 (N.D. Cal. Oct. 30, 2023). Likewise, in *Kadrey v. Meta Platforms, Inc.*, the trial court ruled that the class action plaintiffs' unfair competition claims must also be dismissed. *See* No. 23-CV-03417, 2023 WL 8039640, at *2 (N.D. Cal. Nov. 20, 2023). The court noted that "[t]o the extent it is based on the surviving claim for direct copyright infringement, it is preempted. To the extent it is based on allegations of fraud or unfairness separate from the surviving copyright claim, the plaintiffs have not come close to alleging such fraud or unfairness." *Id.* (citations omitted) (first citing 17 U.S.C. § 301(a); and then citing Maloney v. T3Media, Inc., 853 F.3d 1004, 1010 (9th Cir. 2017)).

210 For example, in *Sega Enterprises Ltd. v. Accolade, Inc.* and again in *Sony Computer Entertainment, Inc. v. Connectix Corp.*, the Ninth Circuit found that reverse engineering a gaming console in order to produce interoperable games (Sega) and a rival gaming platform (Sony) was fair use. Sega Enters. Ltd. v. Accolade, Inc., 977 F.2d 1510, 1523 (9th Cir. 1992); Sony Comput. Ent., Inc. v. Connectix Corp., 203 F.3d 596, 608 (9th Cir. 2000). In both cases the Ninth Circuit found that there was no cognizable market effect because the rival products did not contain any protectable expression derived from the plaintiffs' consoles. *Sega Enters.*, 977 F.2d at 1523; *Sony Comput. Ent.*, 203 F.3d at 608. The defendants were entitled to use uncopyrightable elements from those consoles to make new independent creative expression possible. *Sega Enters.*, 977 F.2d at 1523; *Sony Comput. Ent.*, 203 F.3d at 608.

211 499 U.S. 340 (1991).

212 *Feist Publications, Inc.*, 499 U.S. at 349-50 (citations omitted) (first quoting Harper & Row Publishers, Inc. v. Nation Enters., 471 U.S. 539, 589 (1985) (Brennan, J., dissenting); then quoting *id.*; and then quoting U.S. Const. art. I, § 8, cl. 8).

213 Andy Warhol Found. for Visual Arts v. Goldsmith, 143 S. Ct. 1258, 1276 (2023) ("[T]he first factor also relates to the justification for the use. In a broad sense, a use that has a distinct purpose is justified because it furthers the goal of copyright, namely, to promote the progress of science and the arts, without diminishing the incentive to create.").

214 *Id.* at 1277 ("An independent justification like this is particularly relevant to assessing fair use where an original work and copying use share the same or highly similar purposes, or where wide dissemination of a secondary work would otherwise run the risk of substitution for the original or licensed derivatives of it.").

215 *See id.*

216 *See* 17 U.S.C. § 107(4).

217 Sag, *supra* note 33, at 338-43.

218 *See id.* at 143-59.

219 804 F.3d 202, 222-23 (2d Cir. 2015) (explaining the ways in which Google had "constructed the snippet feature in a manner that substantially protects against its serving as an effectively competing substitute for Plaintiffs' books").

220 Letter from Fred von Lohmann, Assoc. Gen. Couns., Copyright, OpenAI, to Shira L. Perlmutter, Reg. of Copyrights & Dir. of the U.S. Copyright Off., U.S. Libr. of Cong. (Oct. 30, 2023), https://www.regulations.gov/comment/COLC-2023-0006-8906 [https://perma.cc/9DXT-T39M] ("OpenAI has employed numerous measures to reduce the incidence of [verbatim repetition or memorization of training data], and we regularly update our practices to deploy more."); *see also* Alex Nichol, *DALL•E 2 Pre-Training Mitigations*, OpenAI Research (June 28, 2022), https://openai.com/research/dall-e-2-pre-training-mitigations [https://perma.cc/9TMH-UVKV].

221 On inducement generally, see MGM Studios, Inc. v. Grokster, Ltd., 545 U.S. 913 (2005).

92 FDMLR 1887

End of Document