How Generative AI Turns Copyright Upside Down[1]

Mark A. Lemley[2]

Abstract

While courts are litigating many copyright issues involving generative AI, from who owns AI-generated works to the fair use of training to infringement by AI outputs, the most fundamental changes generative AI will bring to copyright law don't fit in any of those categories. The new model of creativity generative AI brings puts considerable strain on copyright's two most fundamental legal doctrines: the idea-expression dichotomy and the substantial similarity test for infringement. Increasingly creativity will be lodged in asking the right questions, not in creating the answers. Asking questions may sometimes be creative, but the AI does the bulk of the work that copyright traditionally exists to reward, and that work will not be protected. That inverts what copyright law now prizes. And because asking the questions will be the basis for copyrightability, similarity of expression in the answers will no longer be of much use in proving the fact of copying of the questions. That means we may need to throw out our test for infringement, or at least apply it in fundamentally different ways.

Generative artificial intelligence (AI)[3] will require us to fundamentally change how we think about creativity, and as a result, how we approach copyright law. The difference isn't simply that AI changes the economics of creativity, making a literally endless array of new content available for essentially no cost. That's true, and I and others have written elsewhere about the market disruptions it may cause.[4] Nor is it the challenge that training AI models presents to copyright liability and fair use doctrines. Those issues are significant, and they have already begun to be litigated.[5] I have argued elsewhere that copyright law correctly permits the training of foundation models like chatGPT.[6] The issues of liability for the occasional AI output that is similar to a particular copyrighted work are somewhat harder.[7] But they aren't fundamental to the

---

[3] In this paper I use the terms "generative AI" and "foundation models" interchangeably to refer to AI systems that train on large general data sets and are capable of producing creative output in multiple domains. Large language models (LLMs) like chatGPT are a particular type of foundation model focused specifically on text.

[4] *E.g.*, Mark A. Lemley, *IP in a World Without Scarcity*, 90 N.Y.U. L. REV. 461, 506 (2018) ("Market disruption will come in fits and starts . . . [b]ut in a post-scarcity world, high-cost products will increasingly become the exception, not the norm."); Deven R. Desai & Mark A. Lemley, Editorial, *Scarcity, Regulation, and the Abundance Society*, FRONTIERS (Jan. 25, 2023), https://www.frontiersin.org/articles/10.3389/frma.2022.1104460/full ("Abundance lowers costs . . . [w]hen that happens, the elimination of scarcity changes the economics of how goods and services are produced and distributed."); Dan L. Burk, *Cheap Creativity and What It Will Do*, 57 GA. L. REV. 1669, 1680 (2023) ("AI threatens to lower the costs of creation itself, so that creative works—however easy to access—are cheap to produce in the first place.").

[5] *E.g.*, DOE 1 v. GitHub, Inc., No. 4:22-cv-06823 (N.D. Cal. filed Nov. 3, 2022); Andersen v. Stability AI Ltd., No. 3:23-cv-00201 (N.D. Cal. filed Jan. 13, 2023). Full disclosure: I represent Stability AI in defending those lawsuits. My opinions here are my own, not my client's.

[6] Mark A. Lemley & Bryan Casey, *Fair Learning*, 99 TEX. L. REV. 743, 748 (2021) ("In this Article, we argue that ML systems should generally be able to use databases for training, whether or not the contents of that database are copyrighted."); Matthew Sag, *The New Legal Landscape for Text Mining and Machine Learning*, 66 J. COPYRIGHT OFC. SOC'Y 291 (2019); Matthew Sag, *Copyright and Copy-Reliant Technology*, 103 NW. U. L. REV. (2009).

[7] *See* Peter Henderson, Xuechen Li, Dan Jurafsky, Tatsunori Hashimoto, Mark A. Lemley & Percy Liang, *Foundation Models and Fair Use*, __ **J. Mach. Learning R.** __ (forthcoming 2024), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4404340 ("These models are usually

new technology, for the simple reason that generative AI is not about copying existing works but about creating new ones; the circumstances in which the output is similar to any given input will be rare.[8]

A number of scholars have begun to consider a third question: who, if anyone, owns those AI-created works.[9] They have come to a variety of answers. Some deny there is a new issue at all, treating AI as a tool like Microsoft Word or Photoshop.[10] Some suggest that AI itself should be an author,[11] though U.S. copyright law currently

---

capable of generating content similar to copyrighted data, and deploying them can potentially impact economic markets that benefit the original data creators."); Matthew Sag, *Copyright Safety for Generative AI*, 61 HOUS. L. REV. (2023).

[8] The output copying issue is complicated by the way the models are designed and the particular ways users may try to prompt models to infringe. I discuss those issues in more detail in *id*.

[9] *E.g.*, Dan L. Burk, *Thirty-Six Views of Copyright Authorship, By Jackson Pollock*, 58 HOUS. L. REV. 263, 266 (2020); Ryan Benjamin Abbott & Elizabeth Rothman, *Disrupting Creativity: Copyright Law in the Age of Generative Artificial Intelligence*, FLA. L. REV. (forthcoming) (manuscript at 42) (on file at https://ssrn.com/abstract=4185327); Andres Guadamuz, *Artificial Intelligence and Copyright*, WIPO MAGAZINE, Oct. 2017, https://www.wipo.int/wipo_magazine/en/2017/05/article_0003; CONGRESSIONAL RESEARCH SERVICE, GENERATIVE ARTIFICIAL INTELLIGENCE AND COPYRIGHT LAW 3 (2023), https://crsreports.congress.gov/product/pdf/LSB/LSB10922

[10] *E.g.*, Dan L. Burk, *Thirty-Six Views of Copyright Authorship, By Jackson Pollock*, 58 HOUS. L. REV. 263, 266 (2020) ("There is never any question of assigning authorship to a paintbrush, saxophone, or word processor, despite their direct involvement in the act of expressive creation."); Michael D. Murray, *Tools Do Not Create: Human Authorship in the Use of Generative Artificial Intelligence* (working paper 2023) (arguing that generative AI does nothing creative and that the creative process for art is similar whether or not the artist uses AI). While that is certainly true of some simple technological uses, I think Burk's argument understates what AI is already capable of and is likely to be even less sustainable as the technology develops. And Murray's argument does not reflect the way the technology actually operates today. But in any event, Burk's view doesn't make the creativity issue go away; it just requires him to pick a human actor that is most closely connected to the creative work and treat that human as the author.

[11] *E.g.*, Ryan Benjamin Abbott & Elizabeth Rothman, *Disrupting Creativity: Copyright Law in the Age of Generative Artificial Intelligence*, FLA. L. REV. (forthcoming) (manuscript at 42) (on file at https://ssrn.com/abstract=4185327) ("Once the desirability of protecting these works is acknowledged, acknowledging AI authorship then becomes nothing more than opting for

rejects that position,[12] or, more plausibly, that the company that owns or deploys the AI

is the author and owner of whatever it creates.[13] U.K. law, one of the few to protect

computer-generated works, declares the author to be "the person by whom the

arrangements necessary for the creation of the work are undertaken – presumably the

reality instead of elaborate legal fictions."); Emmanuel Salami, *AI-Generated Works and Copyright Law: Towards a Union of Strange Bedfellows*, 16 J. INTELL. PROP. L. & PRAC. 124, 124 (2020) ("This article proposes for the ascription of legal personhood to AI systems to ensure that AI is recognized as an author of its works and creations under copyright law."). *But see* James Grimmelmann, *There's No Such Thing as a Computer-Authored Work – And It's a Good Thing, Too*, 39 COLUM. J. L. & ARTS 403, 403 (2016) ("Copyright law doesn't recognize computer programs as authors, and it shouldn't."). *Cf.* Mala Chatterjee & Jeanne C. Fromer, *Minds, Machines, and the Law: The Case of Volition in Copyright Law*, 119 COLUM. L. REV. 1887 (2019) (discussing machine volition in the context of infringement, not authorship).

[12] Naruto v. Slater, 888 F.3d 418, 426 (9th Cir. 2018) (holding that only a human can be an author); U.S. COPYRIGHT OFFICE, COMPENDIUM OF U.S. COPYRIGHT OFFICE PRACTICES § 306 (3d ed. 2021) ("The U.S. Copyright Office will register an original work of authorship, provided that the work was created by a human being."); Pamela Samuelson, *Allocating Ownership in Computer-Generated Works*, 47 U. Pitt. L. Rev. 1185, 1199, 1192 (1986) ("it is still fair to say that it was not withing Congress' contemplation to grant intellectual property rights to machines."). The current law is a bit hard to square with the law's treatment of corporations, which Charles Stross has called the original AIs, as authors in their own right. Charles Stross, Keynote Address at the 34th Chaos Communication Congress (Dec. 2017), http://www.antipope.org/charlie/blog-static/2018/01/dude-you-broke-the-future.html ("I'm talking about the very old, very slow AIs we call corporations, of course."). *See also* U.S. COPYRIGHT OFFICE, CIRCULAR 30: WORKS MADE FOR HIRE 4 (explaining that as defined in 17 U.S.C. § 101 "[i]f a work is a work made for hire, the employer or the party that specially ordered or commissioned that work is the author of that work."). If corporations can be not just owners but authors, it's not clear why other artificial entities couldn't be.

[13] *E.g.*, Andres Guadamuz, *Artificial Intelligence and Copyright*, WIPO MAGAZINE, Oct. 2017, https://www.wipo.int/wipo_magazine/en/2017/05/article_0003.html ("Granting copyright to the person who made the operation of artificial intelligence possible seems to be the most sensible approach"); *see also* CONGRESSIONAL RESEARCH SERVICE, GENERATIVE ARTIFICIAL INTELLIGENCE AND COPYRIGHT LAW 3 (2023), https://crsreports.congress.gov/product/pdf/LSB/LSB10922 ("The creative choices involved in coding and training the AI, on the other hand, might give an AI's creator a stronger claim to some form of authorship than the manufacturer of a camera [who does not retain copyright in photographs taken by the cameras they produce]."). Notably, the current generation of AI companies does not retain copyright in the output of its works. *E.g.*, Terms of Use, OPENAI, https://openai.com/policies/terms-of-use (last updated Mar. 14, 2023) ("OpenAI hereby assigns to you all its right, title and interest in and to Output."); Terms of Use, STABILITY AI, https://stability.ai/terms-of-use (last updated Dec. 20, 2022).

user of the AI.[14] Others have suggested that no one owns AI-generated creative works

and that they fall into the public domain.[15] That would also seem to be the consequence

of Shyam Balganesh's theory that human causation is the sine qua non of copyright.[16]

---

[14] https://www.gov.uk/government/consultations/artificial-intelligence-and-ip-copyright-and-patents/artificial-intelligence-and-intellectual-property-copyright-and-patents. China takes the same approach. Li v. Liu, (2023) Jing 0491 Min Chu No. 11279 (Beijing Internet Court) (English translation on file with author). *See also* Annemarie Bridy, *Coding Creativity: Copyright and the Artificially Intelligent Author*, 2012 STAN. TECH. L. REV. 1, 20 ("artificially intelligent computer programs that autonomously generate art need not be relegated for copyright purposes to scare-quoted authorship; their works can be regarded as proper works of authorship . . . by virtue of their nexus to human creativity."); Edward Lee, *Prompting Progress: Authorship in the Age of AI* (working paper 2023). Who that is in the context of generative AI (the writer of the algorithm? The trainer? The owner of the computer it runs on? The person who prompts the AI?) is far from clear. *Cf.* Bryan Casey & Mark A. Lemley, *Remedies for Robots*, 86 U. CHI. L. REV. 1311 (2020) (noting a similar problem in assigning liability for AI error).

Interestingly, the U.K. takes the opposite approach with patents, denying protection for AI-generated inventions. Thaler v. Comptroller-General of Patents, Designs and Trademarks, [2023] UKSC 49.

[15] *E.g.*, Matt Blaszczyk, *Impossibility of Emergent Works' Protection in U.S. and EU Copyright Law,* 25 N.C. J. L. & Tech. 1 (2023); Grimmelmann, *supra* note __ (discussing the public domain as a possible outcome); Jane C. Ginsburg & Luke Ali Budiardja, *Authors and Machines*, 34 BERKELEY TECH. L.J. 343, 446-47 (2019) (arguing that where "no human participant would meet the requirements of authorship . . . the work is not a work of authorship and thus falls outside of copyright's domain"); Haochen Sun, *Redesigning Copyright Protection in the Era of Artificial Intelligence*, 107 IOWA L. REV. 1213, 1217 (2022) ("However, AI works generated solely by autonomous AI systems should be placed in the public domain without copyright protection."); Michael Kasdan & Brian Pattengale, *A Look at Future AI Questions for the US Copyright Office*, LAW360 (Nov. 10, 2022), https://g2bswiggins.wpenginepowered.com/wp-content/uploads/2022/11/Law360-A-Look-At-Future-AI-Questions-For-The-US-Copyright-Office.pdf#page=7 ("If recent case law from the patent law side of the aisle is to be any guide, it is likely that courts will continue to find that AI will not be eligible as the author of a work. Therefore, any such AI-generated works would be public domain.").

The phrase "fall into the public domain" was coined by nineteenth century poet Alfred de Vigny, who analogized a work's expiration of copyright protection with it falling "into the sink hole of public domain." PAUL TORREMANS, COPYRIGHT LAW: A HANDBOOK OF CONTEMPORARY RESEARCH 154 (2007). It's an odd phrase, as if the expiration of copyright protection were some sort of accident or step down. It isn't. The public domain is a central part of the copyright balance. *See* Benjamin Ely Marks, *Copyright Protection, Privacy Rights, and the Fair Use Doctrine: The Post-*Salinger *Decade Reconsidered*, 72 N.Y.U L. REV. 1376, 1377 ("The control granted by the federal copyright statute, however, is neither perpetual nor absolute . . . [r]ather, the author's property interest in his or her intellectual or artistic production is balanced against the public's interest in free and immediate access to materials essential to the

There may be a middle ground that centers copyright around whether enough human creativity goes into structuring the prompt. The Copyright Office has taken that position in theory,[17] though in practice they have rejected copyright registrations even when the user engaged in significant prompt engineering.[18] On this "prompt-based" approach, we ignore the creativity contributed by the AI but continue to reward creativity contributed by users, assuming the prompt or series of prompts is detailed enough to rise to the level of creative choice.[19] That seems to be the way practice is trending, particularly in light of current Copyright Office guidance. And as I note below, as AI-generated works become more valuable there will be strong incentives to find someone who owns it.

---

development of society."); WIPO Copyright Treaty, Dec. 20, 1996, 36 I.L.M. 65 (1997) (emphasizing the need to "maintain a balance between the rights of authors and the larger public interest, particularly education, research and access to information"). *But cf.* Anupam Chander & Madhavi Sunder, *The Romance of the Public Domain*, CALIF. L. REV. 1331, 1334 (2004) ("But we are also concerned that the increasingly binary intellectual property debates--in which we must choose either intellectual property or the public domain–obscures other important interests, options, critiques, and claims for justice that are embedded in property rights."). Perhaps we should speak of works "rising into the public domain" instead.

[16] Shyamkrishna Balganesh, *Causing Copyright*, 117 Colum. L. Rev. 1, 18-23 (2017) (discussing the role of mechanical intervention in copyright protection and postulating, but not resolving, the prospect of AI creativity).

[17] *See* Copyright Office Statement of Policy, 88 Fed. Reg. 16192 (Mar. 16, 2023) ("In the case of works containing AI-generated material, the Office will consider whether the AI contributions are the result of 'mechanical reproduction' or instead of an author's 'own original mental conception, to which [the author] gave visible form.' The answer will depend on the circumstances, particularly how the AI tool operates and how it was used to create the final work.")).

[18] https://fingfx.thomsonreuters.com/gfx/legaldocs/byprrqkqxpe/AI%20COPYRIGHT%20REGISTRATION%20decision.pdf. To be sure, the Copyright Office's ruling depended on the artist's unwillingness to disclaim the parts of the work created by AI.

[19] Balganesh would presumably endorse this approach, since it looks for human agency in creation.

Our current legal doctrine is not well designed to support a prompt-based copyright system, however, as I argue in this paper. Such a world requires us to upend our two most fundamental legal doctrines, the idea-expression dichotomy that governs protectability and the substantial similarity test for copyright infringement. If copyrightability exists at all for AI-generated works, it will be thin, and the things that make AI-generated output attractive or valuable may well be things copyright doesn't protect. And proving infringement will be much harder than it is today, because our most powerful tool for showing copying – probative similarity – is of little or no use when it comes to generative AI. Alternatively, if we decide (as the Copyright Office so far has) not to protect AI-generated work at all, copyright law itself will slide into irrelevance as more and more creativity is bound up with AI.

I.   Idea-Expression

The idea-expression dichotomy is the most fundamental limit on the scope of copyright protection. Copyright protection does not extend to ideas or functional elements in a work, but only to the creator's particular way of expressing those ideas.[20] This long-standing principle is designed to ensure that copyright owners don't end up

---

[20]   17 U.S.C. § 102(b) ("In no case does copyright protection for an original work of authorship extend to any idea . . . regardless of the form in which it is described, explained, illustrated, or embodied in such work."); *see also* Baker v. Selden, 101 U.S. 99, 105 (1879) ("The description of the art in a book, though entitled to the benefit of copyright, lays no foundation for an exclusive claim to the art itself."). Section 102(b) is broader than the idea-expression dichotomy, as Pam Samuelson has noted. Pamela Samuelson, *Why Copyright Law Excludes Systems and Processes from the Scope of Its Protection*, 85 TEX. L. REV. 1921 (2007). Those other limits will come into play when generative AI creates computer code. But for purposes of LLMs, the idea-expression dichotomy is the critical part.

controlling an entire concept or genre, but only the particular way they have chosen to communicate that idea. The idea itself is free for the world to use.

Ensuring that copyright doesn't protect ideas is not just embedded in the caselaw and the Copyright Act. It is a fundamental part of what makes copyright law constitutional. Along with the fair use defense, the Supreme Court tells us, the idea-expression dichotomy prevents copyright from treading on the First Amendment.[21] And it ensures that copyright "promotes the progress of science" by letting people learn and reuse the ideas expressed. Indeed, if there are only a limited number of ways to express a particular idea, copyright denies all protection rather than risk fencing the idea in.[22]

---

[21] Golan v. Holder, 565 U.S. 302, 329 (2012) ("[The] idea/expression dichotomy strike[s] a definitional balance between the First Amendment and the Copyright Act by permitting free communication of facts while still protecting an author's expression." (quoting Harper & Row, Publishers, Inc. v. Nation Enterprises, 471 U.S. 539, 556 (1985))); Eldred v. Ashcroft, 537 U.S. 186, 219–20 (2003) ("[C]opyright law contains built-in First Amendment accommodations. First, it distinguishes between ideas and expression and makes only the latter eligible for copyright protection.").

It is not at all clear that these guardrails suffice to insulate copyright from today's more robust First Amendment. *See* Mark A. Lemley & Eugene Volokh, *Freedom of Speech and Injunctions in Intellectual Property Cases*, 48 DUKE L. J. 147, 150 (1998) ("In copyright cases . . . preliminary injunctions are granted pretty much as a matter of course, even when the defendant has engaged in creative adaptation, not just literal copying . . . [b]ut libel law and obscenity law are likewise constitutionally valid restrictions on speech, and yet courts refuse to allow preliminary injunctions there."); *cf.* NEIL WEINSTOCK NETANEL, COPYRIGHT'S PARADOX, 4 (2008) ("Copyright is thus a potential impediment to free expression . . . . [I]t may . . . prevent speakers from effectively conveying their message and challenging prevailing views."). That is particularly true after *Andy Warhol Foundation v. Goldsmith*, 598 U.S. __ (2023), where the Supreme Court cut back on the scope of the fair use doctrine as it applied to transformative works. The Court's move to limit fair use for expressive works makes the idea-expression limit all the more central to copyright's balance.

[22] Morrissey v. Procter & Gamble Co., 379 F.2d 675, 678 (1st Cir. 1967) ("When the topic necessarily requires . . . if not only one form of expression, at best only a limited number, to permit copyright would mean that a party or parties, by copyrighting a mere handful of forms, could exhaust all possibilities of future use of the substance.").

That system worked well in a world where the expression of ideas was hard, and was therefore the thing copyright needed to encourage. No one should be entitled to own the idea of a painting of a comet appearing over the beach at sunset. But everyone is free to express that idea in their own way, and the variations in how they express the basic idea is the stuff of copyright. The hard work of creation – actually painting the brushstrokes, making the individual choices of color, shading and composition that bring the painting to life – the artist gets to own.[23]

AI automates much of that formerly hard work. I can ask Dall-E or Midjourney to generate a painting of a comet appearing over the beach at sunset and it will take care of the rest in seconds. If I want more specificity, I can ask it to vary what is in the background, or the colors, or to render it in a particular style, and it will adapt the painting accordingly. I have to come up with the basic idea and tell it what I want. But the AI does the bulk of the work that copyright traditionally exists to reward. Dall-E 3 will even turn your basic request into a better prompt in order to give you better images.[24]

This isn't just true of paintings. ChatGPT writes text for you. It will even develop plot devices and characters to fit the parameters of a story you give it. Song AIs are taking off, some generating AI covers of existing songs but others writing entirely new songs that feel and sound like the songs of existing artists – or even ones

---

[23] OK, not *all* the hard stuff. Coming up with ideas can be hard too. But the fact that creators can protect their expression of an idea makes it harder to just take the idea wholesale, so it ends up giving them an incentive to develop and express new ideas.

[24] https://www.platformer.news/p/how-to-see-the-future-using-dall.

that feel entirely new. So it's not just the literal elements of the work, but even many of the non-literal elements of the work, that are contributed by the AI.

The result is that, increasingly, the things humans contribute in a collaboration with generative AI will be ideas and high-level concepts. AI will contribute the expression. That turns copyright law on its head. Perhaps it means that the AI should be treated as the copyright author, though as noted above current law doesn't permit that, and none of the AI companies seem to be pushing for copyright ownership in the expression they generate.

Or perhaps it means that *no one* owns the copyright in AI-generated works: prompters don't own the ideas they contribute, and AIs can't own the expression they contribute. That might make sense as a policy matter – it is not obvious that generative AI needs the incentive of copyright law – but it is unlikely to be politically sustainable. The instinct Rochelle Dreyfuss has derided as "if value then right"[25] is quite strong, and I suspect that as the works AI generates become more valuable people and companies will rush in to claim ownership of them.

The most common and most plausible answer, I think, seeks to locate creativity not in the generation of outputs from AI but in the structuring of the prompts that produce those outputs.[26] On this theory, Midjourney or Dall-E might generate an

---

[25]   *See* Rochelle Cooper Dreyfuss, *Expressive Genericity: Trademarks as Language in the Pepsi Generation*, 65 NOTRE DAME L. REV. 397, 405 (1990); Alfred C. Yen, *Brief Thoughts About If Value/Then Right*, B.U. L. REV. 2479, 2780 (2019) ("In this brief Essay, I will discuss something of interest to Professor Gordon and others, namely the "if value/then right" principle . . . [which] concerns Professor Gordon because she believes that it expresses socially unproductive hostility to free riding.").

[26]   *See, e.g.*, Congressional Research Service, *supra* note 12, at 3 ("[T]he AI user who prompts the creation of a specific work might be compared to the photographer who uses . . . [a] camera to

10

original work of art, but it does so in response to specific instructions from the user. While some instructions are simple, often those instructions are iterated, with users refining the questions they ask chatGPT or the directions they give image generation AIs until they get the output they are looking for.[27] Sometimes they get it right the first time, but users will sometimes go through dozens of refinements to coax an AI to produce just the right output. And there is strong evidence that the way the prompt is phrased or refined can fundamentally influence the output of foundation models. Indeed, companies are now hiring "prompt engineers" with the skill of asking the right

---

capture a specific image. On this view, the AI user would be considered the author and, therefore, the initial copyright owner.").

[27]   Indeed, one putative copyright owner refined their prompt 600 times, producing this work:

Nonetheless, the Copyright Office refused registration, in part because the author would not disclaim the portions of the work attributable to the AI. https://www.copyright.gov/rulings-filings/review-board/docs/Theatre-Dopera-Spatial.pdf.

questions.[28] And "jailbreaking" limits on chatbots by asking questions the right way

has become a sport in Silicon Valley.[29]

The Copyright Office has taken the position that an original work created solely

by a generative AI isn't copyrightable,[30] but has hinted that a user owns the copyright,

---

[28] *E.g.*, Prompt Engineer & Librarian, ANTHROPIC, https://jobs.lever.co/Anthropic/e3cde481-d446-460f-b576-93cab67bd1ed; AI Prompt Engineer, BOOZ ALLEN HAMILTON, https://www.theladders.com/job/ai-prompt-engineer-boozallenhamilton-bethesda-md_63140467. *See generally* Craig S. Smith, *Mom, Dad, I Want to Be a Prompt Engineer*, FORBES (Apr. 5, 2023), https://www.forbes.com/sites/craigsmith/2023/04/05/mom-dad-i-want-to-be-a-prompt-engineer/?sh=9c5f21759c8e; Henderson et al., *supra* note __ (demonstrating that phrasing prompts in certain ways can trigger infringement in outputs).

[29] https://klse.i3investor.com/web/blog/detail/kianweiaritcles/2023-04-09-story-h-274132588-Jailbreaking_AI_chatbots_is_tech_s_new_pastime. My favorite example:

[30] Thaler v. Perlmutter, __ F. Supp. 3d __ (D.D.C. Aug. 18, 2023) (affirming Copyright Office's rejection of an application to register a work created entirely by an AI even though the work

at least in theory, if they give sufficiently detailed instructions to the AI that help shape

the ultimate work.[31] The mere fact that copyrightability turns on iterated instructions

would clearly qualify for copyright protection if made by a human). This is the image held

uncopyrightable:

[31] Statement of Policy, 88 Fed. Reg. 16192 (Mar. 16, 2023) (noting that "when an AI technology receives solely a prompt from a human . . . the 'traditional elements of authorship' are determined and executed by the technology—not the human user" but that "[i]n the case of works containing AI-generated material, the Office will consider whether the AI contributions are the result of 'mechanical reproduction' or instead of an author's 'own original mental conception, to which [the author] gave visible form'").

In practice, however, the Copyright Office has rejected AI-generated works even when they use hundreds of prompts unless the artist disclaims the parts of the work that originate with the AI. https://fingfx.thomsonreuters.com/gfx/legaldocs/byprrqkqxpe/AI%20COPYRIGHT%20REGISTRATION%20decision.pdf. And in the *Zarya of the Dawn* case it has limited copyright in hybrid works generated by both humans and AI to the portions that were unambiguously created by humans (here, the text but not the images were protectable). https://www.copyright.gov/docs/zarya-of-the-dawn.pdf.

gives users an incentive to issue those instructions, because doing so will be the difference between owning the output and having it fall into the public domain. And as people become more accustomed to using generative AI, perhaps they will write more and more detailed prompts to tailor the output to what they want. And a sufficiently detailed direction to a computer may embody creativity, just as a sufficiently detailed instruction to a human camera operator would.[32] The art world already has analogous
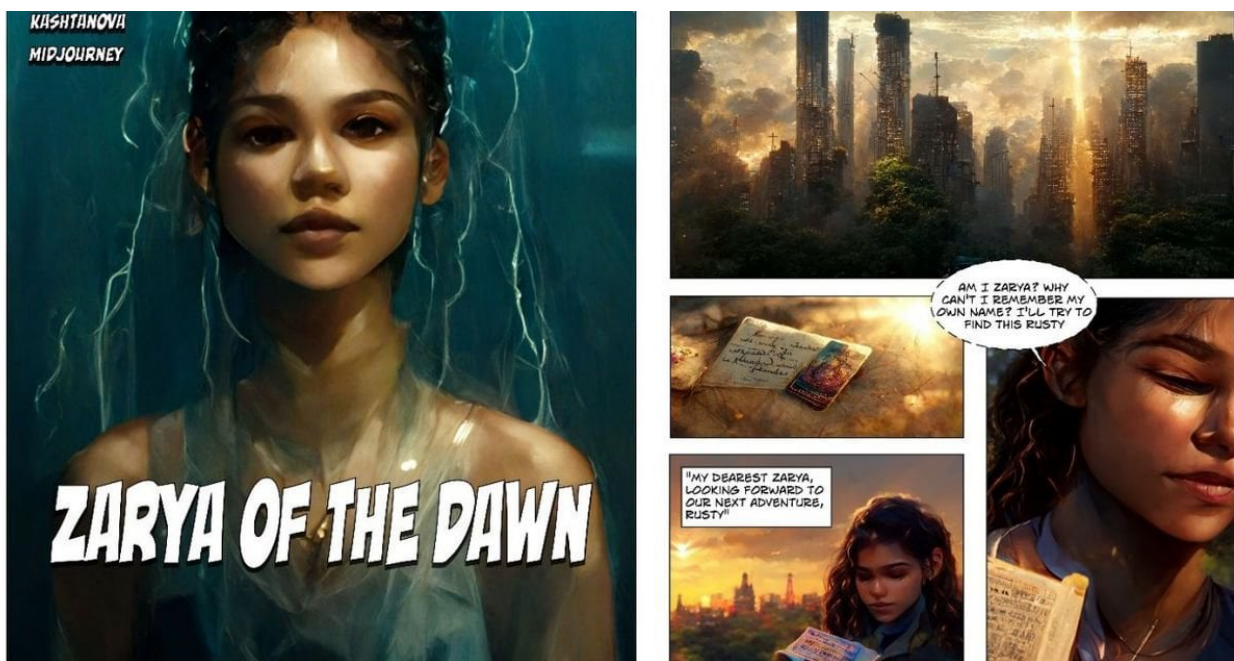
It did the same in the Suryast case, where the artist used their own photograph (one the Copyright Office conceded was itself copyrightable) and asked an AI to render it in the style of Vincent Van Gogh. https://www.copyright.gov/rulings-filings/review-board/docs/SURYAST.pdf. That suggests that the Office may interpret its own policy statement to require execution as well as conception of the expression by a human rather than a machine.

The Copyright Office has put out a detailed request for comments on the issue, so its policy may change. Copyright Office, *Artificial Intelligence and Copyright: Notice of inquiry and request for comments*, No. 2023-6 (Aug. 30, 2023).

[32] Christa Laser suggests photographs as the proper analogy. Christa Laser, *How a Century-Old Insight of Photography Can Inform Legal Questions of AI-Generated Artwork*, https://blog.ericgoldman.org/archives/2023/08/how-a-century-old-insight-of-photography-can-inform-legal-questions-of-ai-generated-artwork-guest-blog-post.htm. *See, e.g.*, Burrow-Giles v. Sarony, 111 US. 53 (1884) (holding that photographs were copyrightable because of the human involvement in setting up the conditions for a photograph); Lindsay v. The Wrecked

14

creativity in the form of artists like Jeff Koons, Dale Chihuly, and Andy Warhol, each of whom came up with the concepts for their works and drew diagrams but delegated the making of the actual artwork to others.[33]

Generative AI makes the actual generation of written work, art, music, and video cheap and easy once you know what you want the AI to make.[34] Human creativity in the world of generative AI, then, will increasingly be about asking the right questions.

As a general matter, I think a limited prompt-based approach makes sense. As we automate the process of generating creative works, the cost and difficulty of producing the actual output declines precipitously.[35] So does the need for copyright. But coming up with the right prompt to generate what you want will sometimes be an art form in itself.[36] And the challenge of asking the right questions can be considerable, particularly if you do it in an iterated fashion until you get the result you want. If you train a generative AI on all the world's music, asking it to give you a new pop song may give you any of a wide range of new songs. Asking it to give you a new pop song that

---

and Abandoned Vessel R.M.S. Titanic, 52 U.S.P.Q.2d 1609 (S.D.N.Y. 1989) (**"**Lindsay's alleged storyboards and the specific directions he provided to the film crew regarding the use of the lightowers and the angles from which to shoot the wreck all indicate that the final footage would indeed be the product of Lindsay's 'original intellectual conceptions.' The fact that Lindsay did not literally perform the filming, i.e. by diving to the wreck and operating the cameras, will not defeat his claims of having "authored" the illuminated footage.")**.** Balganesh collects cases in which those who set up photographs or other displays have been held as authors. Balganesh, *supra* note __, at 18-22.

33 I am indebted to Matt Sag for this point.

34 Burk, *Cheap Creativity*, *supra* note __, at 5.

35 *See* Burk, *supra* note 1, at 1680 ("Thus, rather than lowering the cost of appropriability for works that are expensive to create, AIs make initial creation itself inexpensive.").

36 *See* Lee, *supra* note __. Lee also notes that authorship may arise in the selection of elements of a work, though presumably he would not extend that to, say, selecting one of the four images Midjourney or Dall-E offers in response to a prompt.

15

might have been created if Drake collaborated with The Weeknd will give you something much more specific.[37] But if you want to generate a new song in your own style with a certain mood and particular lyrics, you will be putting significant creative effort into writing the prompt. It seems reasonable for copyright law to encourage that creative effort in the cases in which it happens. And because copyright law requires only a bare minimum of originality, asking more than the most simple question may be enough to qualify for copyright protection.[38]

But if it does, the resulting copyright protection is going to be extremely narrow. It will not extend to the ideas for a painting or story, to underlying facts, or to functional elements of a prompt. And it won't extend to the AI's expression of those ideas or functional elements.[39] Any resulting protection will have to triangulate somewhere between the two, finding creativity in some subset of prompt instructions that are sufficiently detailed to be neither ideas nor dictated by the ideas. And as generative AI

---

[37] And quite good, as it happens. https://mashable.com/article/drake-weeknd-ai-song-controversy. But you're unlikely to be able to hear it by the time this article comes out; it disappeared almost immediately as a result of (somewhat dubious) copyright takedown notices. You can, however, find the song without their voices – something that is entirely generated by AI. https://www.youtube.com/watch?v=_0y0JDwJhZg.

[38] *See* Alfred Bell & Co. v. Catalda Fine Arts, Inc., 191 F.2d 99, 102 (2d Cir. 1951) ("'Original' in reference to a copyrighted work means that the particular work 'owes its origin' to the 'author.' No large measure of novelty is necessary."); The Copyright Office will not register copyrights on short words and phrases, though at least one court has held (wrongly, I think) that six words is enough for copyright protection. Hall v. Swift, 786 F. App'x 711, 712 (9th Cir. 2019).

[39]

https://fingfx.thomsonreuters.com/gfx/legaldocs/byprrqkqxpe/AI%20COPYRIGHT%20REGISTRATION%20decision.pdf (refusing copyright registration despite hundreds of prompts because the Office believed Midjourney, not the artist, was ultimately responsible for the work). It is notable that the artist did not disclose the prompt he ultimately used and did not claim copyright in the prompt itself.

increasingly rewrites the prompts for you, even many aspects of the prompts may not be copyrightable. The thing copyright will protect is not the core expression of the work, but a few peripheral elements on the border between idea and expression, whether in the text of the prompt itself or in extremely narrow strands of creativity in the AI-generated work that can somehow be traced to the creative elements of the underlying prompt. And protection for even those narrow elements may be in doubt, because the random seed included in most generative AI programs means that even careful prompting will not produce a deterministic outcome. Courts and the Copyright Office have rejected protection where the purported author set the conditions for a work to be created but could not control how it actually developed.[40]

This isn't merely Jeff Koons or Dale Chihuly in a computer. Their works are copyrightable because the art was made by humans. *Who* the author is when one human delegates to another is an interesting question, but it doesn't affect the fact that the resulting work is protectable. As long as creativity contributed by the AI doesn't count, by contrast, the thing that we have thought of as the heart of (and the reason for) copyright protection won't qualify for protection in an AI-generated work.

II.   Substantial Similarity.

---

[40]   Kelley v. Chicago Park Dist., 635 F.3d 290 (7th Cir. 2011) (rejecting copyright for a garden because even though it was planted according to a human design, the way the garden actually grew was a function of natural processes beyond the author's control); https://www.copyright.gov/rulings-filings/review-board/docs/Theatre-Dopera-Spatial.pdf (Copyright Office decision citing *Kelley* to deny protection to an AI-generated work despite hundreds of prompts used by the authors).

A prompt-based model of copyright also upends our test for infringement.[41] Our basic system for proving infringement requires a plaintiff to show (1) that the defendant copied from the plaintiff, and (2) that *what* the defendant copied from the plaintiff is a more than de minimis amount of protectable expression.[42] If the fact of copying is disputed, courts traditionally permit a factfinder to infer copying from proof of access coupled with substantial similarity between the works.[43] And in practice, because access at least to published works is ubiquitous in the internet era, the question of copying has come down in practice to evidence of similarity between the works.[44] Indeed, when works are sufficiently similar we often presume copying even in the

---

[41] Robert Heverly makes a different argument – that if AI cannot be an author because it cannot act consciously, it cannot act with volition and therefore cannot infringe a copyright. Robert Heverly, *AI, Creativity, and Liability for Direct Copyright Infringement* (working paper 2023); *cf.* Balganesh, *supra* note __ (requiring human causation). I think the volition issue is an interesting one, though I am not as sure as Heverly that it is directly linked to AI authorship. In any event, it is beyond the scope of this paper.

[42] Arnstein v. Porter, 154 F.2d 464, 468 (2d Cir. 1946) ("In applying that standard here, it is important to avoid confusing two separate elements essential to a plaintiff's case in such a suit: (a) that defendant copied from plaintiff's copyrighted work and (b) that the copying (assuming it to be proved) went to far as to constitute improper appropriation.").

[43] *See id.* ("As to the first—copying—the evidence may consist (a) of defendant's admission that he copied or (b) of circumstantial evidence—usually evidence of access—from which the trier of the facts may reasonably infer copying."); Rentmeester v. Nike, Inc., 883 F.3d 1111, 1124 (9th Cir. 2018) ("The inverse ratio rule provides that the stronger the evidence of access, the less compelling the similarities between the two works need be in order to give rise to an inference of copying . . . [and] the more compelling the similarities supporting an inference of copying, the less compelling the evidence of access need be.").

[44] *See* Skidmore ex rel. Wolfe v. Led Zeppelin, 952 F.3d 1051, 1068 (9th Cir. 2020) ("Given the ubiquity of ways to access media online, from YouTube to subscription services like Netflix and Spotify, access may be established by a trivial showing that the work is available on demand."). *But cf.* Clark D. Asay, *An Empirical Study of Copyright's Substantial Similarity Test*, 13 U.C. IRVINE L. REV. 35, 75 (2022) ("Slightly over 52% of opinions with a decision on prong one used access as the sole subtest for determining prong one's [factual copying] outcome . . . only a little over 25% of the opinions assessed some form of similarity in determining prong one, whether it be striking, probative, substantial, or some other denomination for similarities between the two works.").

absence of a good story for how it happened, deciding that it must have happened subconsciously.[45]

But that system founders in the world of generative AI.

If you and I both ask MidJourney to give us an impressionist-style painting of penguins having a picnic at a beach, we are likely to get, not identical pictures (they are generated anew on each request based on the combination of the model weights, the prompt, and a random seed),[46] but at least somewhat similar ones.[47] But that similarity

---

[45] *See* Bright Tunes Music Corp. v. Harrisongs Music, Ltd., 420 F. Supp. 177, 180 (S.D.N.Y. 1976), aff'd sub nom. ABKCO Music, Inc. v. Harrisongs Music, Ltd., 722 F.2d 988 (2d Cir. 1983) ("As he tried this possibility and that, there came to the surface of his mind a particular combination that pleased him as being one he felt would be appealing to a prospective listener; in other words, that this combination of sounds would work. Why? Because his subconscious knew it already had worked in a song his conscious mind did not remember.").

[46] *See* Katherine Lee et al., *Talkin' About AI Generation: Copyright and the Generative AI Supply Chain* (working paper 2023).

[47] And we might not even get that. For example, here is what MidJourney gave me in response to that request:

doesn't result from any copyrightable expression at all; a request that simple is just an idea and doesn't confer copyrightability at all. It results from two things: the same idea, and the fact that the AI generated expression based on its understanding of the concept

Matt Sag put variants of the same prompt into different AIs, including MidJourney, and got some very different images just by adding the word "impressionist":

we asked it to execute. Something similar is true of chatGPT, which generates text based in its prediction of the next logical token in a sentence. Give it the same seed prompt, and it is likely to generate similar answers, though because the predictive text generation happens piece by piece (using "tokens" that may be as small as a single word or even a syllable), small changes early on can lead to larger divergences.[48] And the more specifically the prompt identifies a particular work in the training dataset, the more likely chatGPT is to replicate it exactly.[49] Those similarities aren't probative of copying of whatever copyrightable elements happen to be in a prompt; they could just as easily result from independent creation of similar prompts or from the commonality of terms already in the training data that do not owe their origin to either party.[50]

---

[48] *See* Maria Diaz, *How to Use ChatGPT*, ZDNet, June 30, 2023, https://www.zdnet.com/article/how-to-use-chatgpt/("Most of the time, when different people ask ChatGPT the same question, they will get the same answer. There might be a few variations in words, but the response will be almost identical.").

[49] Henderson et al., *supra* note __ (showing that a general request to ChatGPT for a story about children attending a wizarding school doesn't produce a story with much similarity to Harry Potter, but that asking for a story that begins with the text of the first paragraph of *Harry Potter and the Sorcerer's Stone* causes it to generate several pages of the book almost verbatim). By contrast, a generic prompt like "write me a poem" is unlikely to generate similar poems each time.

[50] True, if the AI company posts the prompt and the response publicly, it is possible that a third party will copy that output directly from the posted site. That is a more straightforward example of copyright infringement, though it will require proof that there was sufficient originality in the prompt that generated the output. This sort of infringement is easier to detect, though it may be desirable to embed unique watermarks in AI output in case we need to trace its provenance in a disputed case. *See, e.g.*, Maurice Schellekens, *Digital Watermarks As Legal Evidence*, 8 DIGIT. EVIDENCE AND ELEC. SIGNATURE L. REV. 152, 152 (2011) ("A digital watermark is meta- information that can be added to a work such as a picture or a movie . . . that [can] identify the author or rights holder."); Rosemarie F. Jones, *Wet Footprints? Digital Watermarks: A Trail to the Copyright Infringer on the Internet*, 26 PEPPERDINE L. REV 559, 568 (1999) ("Watermarks could also deter counterfeiters from making illegal copies because an imitation would be easily identifiable from the original.").

 Chris Buccafusco claims that there is no such thing as independent creation, and that most cases of independent creation involve subconscious copying. Christopher Buccafusco,

What we really want to know to determine infringement in a prompt-based copyright system is *not* whether the works are similar but whether I copied the details of your prompt from you. If so, and if that prompt had enough creativity, I may be infringing. But if I generated a substantially similar image by asking a different question, then even though the output is similar and used for competitive purposes, there is no copyright infringement.[51] And conversely, even identical copying of a prompt may not generate outputs existing copyright law would treat as substantially similar, as the penguins illustrate.`

Further, even if I generated a similar or even virtually identical image by giving the AI the very same prompt, that doesn't mean I copied the prompt. Prompts are much simpler than the resulting expression, and it is therefore much more plausible

---

There's No Such Thing as Independent Creation, and It's a Good Thing, Too, 64 WM. & MARY L. REV. 1617 (2023). Even if Buccafusco was right about similar works by humans resulting from subconscious copying rather than independent development – and I don't think he is, see Mark A. Lemley, *The Myth of the Sole Inventor*, 110 MICH. L. REV. 709 (2012) (documenting numerous examples of truly independent invention) – similarity in AI outputs can't be attributed to subconscious copying on the part of the AI.

[51]   The narrow nature of the protectable expression created by humans discussed in the last section may well mean that the test for infringement is not "substantial similarity of protected expression" but the higher standard of "virtual identity." Apple Computer, Inc. v. Microsoft Corp., 35 F.3d 1435, 1439 (9th Cir. 1994). Courts use that higher standard when copyright protection is thin, because in such a case similarity between two works might not be particularly probative of copying. But in the upside-down world of generative AI copyright, even virtual identity between two works may not be completely probative of copying, though as noted above more specific prompts are somewhat more likely to generate identical output, and those more specific prompts may be the ones more likely to pass the creativity threshold. Identity of output may simply reflect identity of the prompts without copying of the prompts, or extreme similarity of outputs resulting from copying the ideas and functional elements of a prompt. *See* Diaz, *supra* note __ (noting that sometimes output will be identical if the prompts are identical).

that two people came up with the same or an extremely similar prompt independently. And independent creation even of an identical work is not copyright infringement.[52]

We have largely ignored the fact that independent creation is not infringing in recent copyright cases; essentially all modern cases assume access and turn on whether there is sufficient similarity between the plaintiff's and defendant's works.[53]

We need to invert that in the world of generative AI. Access will have to play a much larger role and similarity a much smaller role in a prompt-based copyright infringement system.[54] That is true for three reasons. First, because AI and not human creativity is generating the vast majority of the expression in outputs, evidence of similarity will be much less probative than it is elsewhere. Unless those similarities can be traced specifically to the use of the same creative criterion in a prompt, they simply

---

[52] *See, e.g.*, Sheldon v. Metro-Goldwyn Pictures Corp., 81 F.2d 49, 54 (2d Cir. 1936) ("[I]f by some magic a man who had never known it were to compose anew Keats's Ode on a Grecian Urn, he would be an 'author,' and, if he copyrighted it, others might not copy that poem, though they might of course copy Keats's."); Ty, Inc. v. GMA Accessories, Inc., 132 F.3d 1167, 1169 (7th Cir. 1997) ("[I]f independent creation results in an identical work, the creator of that work is free to sell it.").

[53] In *Thomson Reuters Ent. Centre GmbH v. Ross Intelligence Inc.*, __ F. Supp. 3d __ (D. Del. Sept. 25, 2023), where the defendant had used AI to generate headnotes for legal decisions that Thomson argued were too similar to its headnotes, the court denied summary judgment to both sides on the question of whether the final headnotes were copied.

[54] *See* Skidmore ex rel. Wolfe v. Led Zeppelin, 952 F.3d 1051, 1068 (9th Cir. 2020) ("Given the ubiquity of ways to access media online, from YouTube to subscription services like Netflix and Spotify, access may be established by a trivial showing that the work is available on demand."). Shyam Balganesh and Peter Menell suggest that *Skidmore* was wrong to dispose of the sliding scale test. Shyamkrishna Balganesh & Peter S. Menell, *Proving Copying*, 64 Wm. & Mary L. Rev. 299 (2022). While I think they are wrong to ignore *Skidmore*'s concern that in a world of ubiquitous access the test had merely reduced the similarity required to find infringement, their argument about the logic of proof of access will have more relevance in the world of generative AI.

aren't probative of copying.[55] The similarity may come from the random seed or from the training of the AI database, neither of which the copyright owner can lay claim to. Indeed, our existing tests for similarity are likely to mislead, because we will base our similarity assessment on things that would be copyrightable elements if created by a human but are not copyrightable because they owe their origin to the AI.

Second, unlike finished works, prompts aren't generally available to the public. So the plausibility of access is much lower in a case alleging prompt copying than it is in other cases. Courts sometimes allow even wildly improbable theories of access to suffice if they believe based on strong similarity that independent creation simply isn't plausible.[56] And the theory of "it's somewhere on the internet" has largely sufficed to meet the burden of proof on access. But courts will need to curb that instinct when it

---

[55] Virtual identity as opposed to similarity may provide some cues, because the more specific the prompt, the more likely chatGPT is to generate the identical output. *See, e.g.*, Apple Computer, 35 F.3d at 1149 ("When the range of protectable and unauthorized expression is narrow, the appropriate standard for illicit copying is virtual identity."); Domingo Cambeiro Pro. Corp. v. Advent, No. 99-17057, 2000 WL 262597 at *3 (9th Cir. 2000) ("Extrinsic analysis revealed that Cambeiro's works were entitled to thin copyright protection at best . . .[t]hus, even on an intrinsic level, Cambeiro would have been obliged to prove "virtually identical copying." (citation omitted)). But even virtual identity doesn't prove copying in the generative AI world. At most, it increases the likelihood that the prompts were the same. But as noted above, it is much easier for prompts to be independently created than for entire works to be.

[56] *See, e.g.*, Ferguson v. National Broadcasting Co., Inc., 584 F.2d 111, 113 (5th Cir. 1978) (where there is striking similarity that precludes the possibility of independent creation, "'copying' may be proved without a showing of access."); Bouchat v. Baltimore Ravens, 41 F.3d 350 (4th Cir. 2000); Ty, Inc. v. GMA Accessories, Inc., 132 F.3d 1167, 1170 (7th Cir. 1997) ("a similarity that is so close as to be highly unlikely to have been an accident of independent creation is evidence of access."); And sometimes even when there isn't much similarity. *See, e.g.*, Arnstein v. Porter, 154 F.2d 464, 469 (2d Cir. 1946) ("The similarities, however, are sufficient so that, if there is enough evidence of access to permit the case to go to the jury, the jury may properly infer that the similarities did not result from coincidence . . . . Although part of plaintiff's testimony on deposition (as to 'stooges' and the like) does seem 'fantastic,' yet plaintiff's credibility, even as to those improbabilities, should be left to the jury."). Arnstein reflects a view of the procedural standards for summary judgment that likely wouldn't survive today.

comes to AI. Absent a plausible story for how the defendant somehow had access to the plaintiff's prompt, even virtual identity doesn't make out a case of copying.[57] And people worried about being sued may be able to establish "cleanroom" procedures to show that they had no access to such a prompt.

If the prompt is copyrightable, that copyrightable material *may* (or may not) end up influencing the expression of the resulting work. If it does, and if the defendant copies the work exactly, they will be liable because they copied everything in the output, so they must have copied whatever small elements in the AI-generated work owe their origin to the creativity in the prompt.

A prompt-based copyright system also complicates the second step of our traditional infringement test – improper appropriation. It is not enough for a defendant to have copied from the plaintiff; it matters *what* they copied. Only the copying of a more than de minimis amount of protectable expression is infringement. Copying of the ideas or other unprotectable elements of a work is permissible.[58] So too is copying of the parts of the work that owe their origin to the machine rather than the person. That means that even if we could use similarities between the works to somehow infer copying, that copying might not be illegal.

Even if the defendant's work isn't identical to the plaintiff's, it is possible that a plaintiff can allege infringement of the prompt by copying of whatever constituent

---

[57] Future technological developments may make access cases easier to prove. If an AI company posts prompts as well as responses online, those prompts will be easier to find and copy. And if an AI starts using autocomplete to suggest prompt inquiries, it may itself inadvertently prompt a user to copy a protectable prompt from another.

[58] *See, e.g.,* Satava v. Lowry, 323 F.3d 805 (9th Cir. 2003).

elements of the AI-generated work that owe their origin to the creativity in the prompt. But doing so will face challenges. Because most of what we traditionally think of as the expression in a song or painting comes from the AI, it won't be protected under current law. So a copyright plaintiff will have to show not just that the resulting works look similar, and that the defendant copied the plaintiff's prompt, but that the similarities in the prompts reflect the things that are expressive about those prompts rather than ideas or other functional elements in the prompts.[59]

Finally, copyright's freedom to create independently arguably extends to "reverse engineering" of prompts – looking at a finished image and trying to get an AI to generate a similar image by writing your own, different prompts.[60] If I start from a final image and come up with my own creative prompts to generate a similar final image, arguably I haven't copied anything copyright law protects, because under current law copyright protects *only* the prompts, and I didn't copy those; I figured out a

---

[59] Nor can plaintiffs get around this requirement by alleging that the AI-generated work is a derivative work, as plaintiffs in AI cases have sometimes alleged. Calling something a derivative work doesn't avoid the same requirement of substantial similarity of protectable expression. Litchfield v. Spielberg, 736 F.2d 1352 (9th Cir. 1984).

[60] *See* Sega Enterprises Ltd. v. Accolade, Inc., 977 F.2d 1510, 1518 (9th Cir. 1992), as amended (Jan. 6, 1993) (holding that Accolade's efforts to create game cartridges that were compatible with Sega's console which involved intermediate copying and reverse engineering of Sega's software were "a fair use of the copyrighted work if such disassembly provides the only means of access to those elements of the code that are not protected by copyright and the copier has a legitimate reason for seeking such"); Sony Computer Entertainment, Inc. v. Connectix Corp., 203 F.3d 596, 608 (9th Cir. 2000) (extending the *Sega* holding beyond just interoperability and allowing Connectix to build software that allowed PlayStation games to run on Apple computers).

way to make a similar result using noninfringing prompts.[61] There isn't law on this, but it seems to follow from the logic of existing rules.

> III.   Implications.

Copyright, then, will look very different in a prompt-based creativity system than it does today. Indeed, two of our most important doctrines – the idea-expression dichotomy and substantial similarity – will be turned on their heads. If copyrightability exists at all for AI-generated works, it will be thin, and the things that make AI-generated output attractive or valuable may well be things copyright doesn't protect. And proving infringement will be much harder than it is today, because our most powerful tool for showing copying – probative similarity – is of little or no use when it comes to generative AI.

In itself, there is nothing wrong with that change. Copyright law is designed to encourage human creativity. It may be that copyright law has – and should have – less influence in a world where much of the creativity we value doesn't come from humans and doesn't require the incentive of a century of copyright protection.[62] But achieving the goal of keeping copyright's principles intact will require new tools. We will have to rethink our basic legal tests as they apply to this new generation of works.

---

[61]   By contrast, if I manage to reverse-engineer the same prompts the plaintiff used, I likely infringe when I use those prompts, because I copied them from the work the plaintiff generated using their copyrighted prompts.

[62]   *Cf.* Lemley, *World Without Scarcity*, *supra* note __ (arguing that copyright will recede in importance as machine-assisted creativity becomes cheaper and easier); Burk, *Cheap Creativity*, *supra* note __.

Alternatively, it may be that generative AI causes us to rethink what copyright protects and why. As I noted above, there will be a strong incentive to change the law so that anything valuable is owned by someone.[63] The "if value, then right" impulse[64] is a strong one, and there will be a lot of money at stake. Powerful players will push hard for someone (read: them) to own the output of generative AI. And they are likely to succeed in creating a new ownership right, just as they have with previous technologies. Indeed, in many industries the very fact that no one owns something will seem foreign and scary.[65] I think we should resist that propertization impulse, but I'm not sure we will.

---

[63] *See supra* notes __-__ and accompanying text.

[64] *See* Dreyfuss, *supra* note __, at 405; Yen, supra note 19, at 2780; Mark A. Lemley, *Property, Intellectual Property, and Free Riding*, 83 TEX. L. REV. 1031, 1033 (2004) ("Courts and commentators adopt-explicitly or implicitly-the economic logic of real property in the context of intellectual property cases . . . they then . . . jump from the idea that intellectual property is property to the idea that the IP owner is entitled to capture the full social value of her right.").

[65] Lemley, *World Without Scarcity*, *supra* note __ (noting that economics is based on ownership as a response to scarcity, and struggles to cope with a world of abundance).

The open source world is an obvious counterexample, but the proprietary software industry struggled for years with how to cope with it. *See, e.g.*, William Henry Gates III, *An Open Letter To Hobbyists*, HOMEBREW COMPUTER CLUB (Feb. 3, 1976), https://digibarn.com/collections/newsletters/homebrew/V2_01/gatesletter.html (arguing that free software cannot succeed and asking "[w]ho can afford to do professional work for nothing?" and "[w]hat hobbyist can put 3-man years into programming, finding all bugs, documenting his product and distribute for free?"); Amy Harmon & John Markoff, *Internal Memo Shows Microsoft Executives' Concern Over Free Software*, N.Y. TIMES (Nov. 3, 1998), https://archive.nytimes.com/www.nytimes.com/library/tech/98/11/biztech/articles/03mem o.html ("A[] [leaked] internal memorandum [from 1998] reflecting the views of some of Microsoft Corp.'s top executives … reveals deep concern about the threat of free software . . . ."); Mitch Wagner, *Is Java SE Open Source Software? The Short Answer Is 'Yes.'*, JAVA MAGAZINE (Nov. 17, 2021), https://blogs.oracle.com/javamagazine/post/java-se-open-source-license (explaining that an open-source version of Java, OpenJDK, only became available in 2007 because "in the 1990s the tech industry wasn't as accepting of open source as it is today … [according to] Donald Smith, senior director of product management for Oracle's Java platform group").

Propertizing AI-generated content would also turn copyright on its head, albeit in different ways. If someone owns AI-generated output, the logical choice is the company that runs the AI itself. But users are likely to want ownership of the results of "their" prompts even if most of the creativity in the output doesn't originate with the user. AI companies may well assign the rights to those users; some already do.[66] But that won't solve the copyright problem. Unless the AI is going to record each prior use and alter the algorithm in real time to preclude generation of similar content in response to a different prompt, what it means to "own" AI output will be quite different than what copyright law means by that term today.

A copyright regime that prevents the actual duplication of the output after it is generated will work fine, but we won't be able to rely on similarity as evidence that you copied from my output any more than we can as evidence that you copied my prompt. We will need either an actual (not virtual) identity standard or some sort of tracking or watermarking system to show that you copied the output from me after it was generated rather than generating your own prompt.[67]

Finally, AI may cause some to argue that we should reevaluate the idea-expression dichotomy or even the whole concept of copyright. If it turns out that

---

[66] *E.g.*, Terms of use, OPENAI, https://openai.com/policies/terms-of-use (last updated Mar. 14, 2023) ("OpenAI hereby assigns to you all its right, title and interest in and to Output."); GitHub Copilot Product Specific Terms, GITHUB, https://github.com/customer-terms/github-copilot-product-specific-terms (last updated June 2023) ("GitHub does not claim any ownership rights in Suggestions. You retain ownership of Your Code.").

[67] Alternatively, we might look for the defendant's paper trail: if the defendant generated the content themselves, they should be able to point to the prompt they used to create that content. But not all defendants will keep records of all such searches.

coming up with sufficiently clear questions to generate good outputs using generative AI is a rare creative skill, there will be pressure to protect that creativity even if it means modifying the idea-expression dichotomy. I am not at all sure that's necessary; people seem to be doing quite well using ChatGPT with no training, and the AIs themselves are evolving to help people generate better prompts – all without the need for copyright protection. Even if we think people need incentives to ask the right questions, protecting creative ideas will put pressure on other doctrines in copyright, because it risks allowing one copyright owner to control too much, particularly after the Supreme Court cut back on the fair use defense in *Warhol*.[68] The original theory for copyrighting photographs was that the artist was entitled to very narrow protection limited to the creative choices early photographers made in lighting and setting up their shots.[69] Today courts have forgotten those constraints, and regularly declare that photographs are entitled to "full" and strong protection[70] even when virtually nothing about the photograph can be traced to authorial choices, as in *Warhol*. Even copyright protection that start out very narrow, as copyright for photography did, may end up mutating, undoing the idea-expression dichotomy and allowing protection for ideas implemented by AIs.

A better approach might be to revel in the fact that we get instantly and for free things that used to cost time and money. We need copyright only if we think we won't

---

[68] *Andy Warhol Found. for Creative Arts v. Goldsmith*, 598 U.S. __ (2023).

[69] *Burrow-Giles*, 111 U.S. at __.

[70] *Warhol*, 598 U.S. at __.

get enough creation without it. That may no longer be a worry in the world of generative AI.[71] It may also be that we will recreate scarcity even in a world of abundance, focusing on a desire for exclusivity or authenticity just as some people prize artisanal food and furniture in a world where it can be cheaply mass-produced.[72] And while authenticity may depend in some measure on IP, it is far more likely to turn on trademark than on copyright law.[73] If we take this approach – the one the Copyright Office has so far been leaning towards – copyright itself may well slide into irrelevance as more and more creativity is bound up with AI. Copyright won't disappear altogether, but it might just not matter for most ordinary creativity ten years from now, just as most creativity on the internet today occurs without the incentive of copyright law.[74]

These are complex social problems, presenting questions about the economics of creation, labor displacement, and even the extent to which creativity is an inherently human activity.[75] The current Hollywood strike is in significant part about fears of

---

[71] To be clear, though, we cannot expect – and would not want – to have *only* machine creativity. Among other problems, generative AI works well only if it is trained on human-generated content. *See* Ilia Shumailov et al., *The Curse of Recursion: Training on Generated Data Makes Models Forget*, https://arxiv.org/abs/2305.17493.

[72] *See, e.g.,* Burk, *Cheap Creativity*, *supra* note __, at 2 (arguing that AI will drive social concern over the value of authenticity, just as it has in other markets that industrialized); Desai & Lemley, *supra* note __ (noting NFTs as an example of artificial recreation of scarcity in a world of abundance); Zahr Said, *Fables of Scarcity in IP*, FRONTIERS IN RESEARCH, https://www.frontiersin.org/articles/10.3389/frma.2022.974154/full (same).

[73] *Cf.* Burk, *Cheap Creativity*, *supra* note __, at 18-24 (noting this fact but also noting the poor fit between trademark law and authenticity in a social sense).

[74] *See* Lemley, *World Without Scarcity*, *supra* note __.

[75] It isn't. It's been observed in animals. *See* Lydia M. Hopper & Andrew W. Torrance, *User Innovation: A Novel Framework for Studying Animal Innovation Within a Comparative Context*, 22 ANIMAL COGNITION 1185, 1186 ("Although animals have not created the diversity and

creators being displaced by AI.[76] I understand and sympathize with those concerns;

they echo those from countless prior generations of employees who faced disruption

from new innovations.[77] But those interests must be balanced against the enormous

promise of cheap creativity for consumers and businesses. "Ban the disruptive

---

complexity of material technologies that humans have … they nonetheless innovate. Innovations by animals have been reported in a range of realms including tool construction and use … the eating and processing of novel foods … and the invention of novel gestures to communicate meaning and maintain social bonds.").

Perhaps AI too is already creative. *See, e.g.*, Annemarie Bridy, *Coding Creativity: Copyright and the Artificially Intelligent Author*, 2012 STAN. TECH. L. REV. 1, 2 ("[A]ll creativity is inherently algorithmic and . . . works produced autonomously by computers are . . . less heterogeneous to . . . their human counterparts . . . than appearances may at first suggest."); *AI Outperforms Humans in Creativity Test*, NEUROSCIENCE NEWS (July 6, 2023), https://neurosciencenews.com/ai-creativity-23585/ ("Artificial Intelligence (AI), specifically GPT-4, was found to match the top 1% of human thinkers on a standard creativity test. The AI application ChatGPT, developed using GPT-4, excelled in fluency and originality in the Torrance Tests of Creative Thinking, a widely recognized tool for assessing creativity.").

But some have argued that only humans can be creative. *See, e.g.*, Sean Dorrance Kelly, *A Philosopher Argues That an AI Can't Be an Artist*, MIT TECH. REV. (Feb. 21, 2019), https://www.technologyreview.com/2019/02/21/239489/a-philosopher-argues-that-an-ai-can-never-be-an-artist/ ("[I]t seems to me, nothing but another human being can properly be understood as a genuinely creative artist. Perhaps AI will someday proceed beyond its computationalist formalism, but that would require a leap that is unimaginable at the moment."); Keith Kirkpatrick, *Can AI Demonstrate Creativity?*, 66 COMMC'NS OF THE ACM 21, 22, February 2023 ("Despite the ability of AI to produce creative outputs based on the attributes of existing works, the process is not the same as a human's creativity, which comes from a combination of real-world experience, emotion, and inspiration."); *Burk, supra* note 1, at 1680 ("My own view is that, lacking any degree of self or situational awareness, they cannot be [creative].").

[76] *See, e.g.*, Vishwam Sankaran, *Hollywood AI Backlash: What Striking Writers and Actors Fear About Tech Replacing Role*, THE INDEPENDENT, July 17, 2023, https://www.independent.co.uk/arts-entertainment/films/news/hollywood-ai-writers-strike-tech-b2376457.html.

[77] Lemley, *World Without Scarcity*, *supra* note ___, at Part IV.

technology" has never been the right answer in the past,[78] and I don't think it is with AI either. If it does nothing else, generative AI makes us rethink what we value in creative work and whether and why we want it to be owned. And that is an important question.

Conclusion

There is an old aphorism that says that if you give an infinite number of monkeys typewriters they will eventually create the works of Shakespeare. In theory, in the copyright case of *Shakespeare v. Infinite Monkeys*, Shakespeare should lose, because the monkeys didn't copy from him. In practice, courts have generally found a way for similarity to trump everything else. But that will need to change in the world of generative AI. Our basic copyright doctrines don't fit generative AI. We will struggle to apply the law to comport with the new realities. And it may suggest that copyright itself is a poor fit for the new world of AI-generated works.

---

[78]   *Id*; Mark A. Lemley & Mark P. McKenna, *Unfair Disruption*, 100 B.U. L. REV. 71 (2020) (noting the resistance incumbents have to disruptive technologies, and arguing that the law shouldn't defer to it).