Harvard Journal of Law & Technology
Volume 38, Number 1 Fall 2024

THE WORK OF COPYRIGHT IN THE AGE OF MACHINE PRODUCTION

*Oren Bracha**

ABSTRACT

The advent of powerful Generative Artificial Intelligence ("GenAI") marks a revolutionary moment in our system of cultural production. The challenge it poses is twofold: (1) Internally, to existing copyright-based business models of cultural production; and (2) Externally, by threating human authors more generally, it raises a deeper set of anxieties concerning livelihood, the inherent value of creativity, and deep innovation. This Article seeks to disentangle these two concerns such that understandable variants of the latter do not result in distortions of the former. Such disentanglement is necessary to ensure that copyright does not become the misguided vehicle for addressing larger cultural anxieties about 'machine creativity' that it is ill-suited to handle. By resorting to and elaborating copyright's fundamental principles, this Article argues that, contrary to conventional wisdom, many claims of broad GenAI copyright infringement should be rejected on foundational grounds concerning the proper domain of copyright's subject matter, rather than on secondary grounds regarding the scope of copyright's protection. This applies to both the upstream and downstream levels of the GenAI production process. With respect to the upstream level, the entire central legal debate is misguided. Both sides in this debate assume that reproduction of copyrighted works, strictly as part of the training process, is infringing, and disagree as to whether it is exempted as fair use. However, reproduction strictly limited for training purposes is not infringing on the more fundamental ground of not using any copyrightable subject matter, without ever needing to reach the fair use question. For purposes of copyright, the mere creation of a copy whose expressive use value will be consumed by no one is an irrelevant physical fact. On the downstream level, the argument that copying of creators' "style" is infringing is subject to a similar analysis.

---

* William C. Conner Chair in Law, The University of Texas School of Law. Special thanks to Aviad Tsherniak and Yoav Volloch. For useful comments and conversations, I would also like to thank Bob Bone, Dan Burk, Christopher Buccafusco, Terry Fisher, Willy Forbath, Kristelia García, John Golden, Patrick Goold, Uri Hacohen, Masahiko Kinoshita, Lee Kovarsky, Edward Lee, Neil Netanel, Blake Reid, Matthew Sag, Larry Sager, Pamela Samuelson, Shani Shisha, David Simon, Benjamin Sobel, Talha Syed, Molly Shaffer Van Houweling, Felix Wu, participants of the University of Texas School of Law Drawing Board workshop, participants of the Chicago IP Colloquium, and participants of the Legal Scholars Roundtable on Artificial Intelligence at Emory University School of Law.

GenAI outputs that replicate a creator's general style rather than any specific work pertains to informational subject matter that has always been outside's copyright domain. Subject matter rules play a dual function in this area. Internally, they ensure that protection does not extend to informational elements that copyright's policy balance requires remain unprotected. Externally, they prevent attempts to weaponize copyright to address genuine GenAI cultural policy concerns that the field is neither designed nor equipped to handle.

TABLE OF CONTENTS

I. INTRODUCTION

> WHEN YOU GET THE DRAGON OUT OF HIS CAVE ON TO THE PLAIN AND IN DAYLIGHT, YOU CAN COUNT HIS TEETH AND CLAWS, AND SEE JUST WHAT IS HIS STRENGTH. BUT TO GET HIM OUT IS ONLY THE FIRST STEP. THE NEXT IS EITHER TO KILL HIM, OR TO TAME HIM AND MAKE HIM A USEFUL ANIMAL.
>
> — OLIVER WENDELL HOLMES JR.[1]

Creative machines have captured much attention and headlines recently.[2] The Artificial Intelligence ("AI") revolution has reached the area of expressive creation and its disruptive effect is rapidly descending on the field. Various Generative Artificial Intelligence technologies

1. Oliver Wendell Holmes, Jr., *The Path of the Law*, 10 HARV. L. REV. 457, 469 (1897).

2. To forestall misunderstanding at the outset: by the term "creative machines," I do not mean to impute to machines a status equivalent to that of human authors. I assume neither that we should treat machines as having the status of recognized moral or legal agents nor that machines should be treated as responsible for their expressive output in the same way that humans are. I use the term simply to refer to the concept of machines capable of generating output that will be treated by some as having similar use value to humanly created expression and hence as being a good substitute for it.

("GenAI")[3] are now capable of generating expressive works of impressive quality.[4] Probably the most known example is ChatGPT; it already produces elaborate, and sometimes striking, text responses to user prompts, including short stories, journalistic reports, and poems.[5] But the technology cuts a much wider swath that is sure to only rapidly increase over time. GenAI is already exhibiting remarkable performance in producing expressive works in a variety of media including image, video, and sound.[6] Given these rapid developments, the output of these systems will likely become significant for any expressive form and media that lends itself to a digital format. As this process unfolds, the turmoil and challenges brought by AI have now spread to the field of expressive creation, sparking high stakes disputes and raising fundamental questions. Because copyright law is our central institutional tool for dispensing cultural policy, many of these challenges are laid at its doorstep. In a hailstorm of legal proceedings that seems to be intensifying by the week, a wide variety of plaintiffs are hurling a broad

3. For an explanation of the term Generative Artificial Intelligence, see *infra* text accompanying notes 44–45. The acronym GenAI should not be confused with Artificial General Intelligence ("AGI"). The latter term is the subject of much speculation about a single, autonomous super-AI that can greatly surpass human capabilities in all or most areas. *See* Ben Goertzel, *Human-Level Artificial General Intelligence and the Possibility of a Technological Singularity: A Reaction to Ray Kurzweil's The Singularity Is Near, and McDermott's Critique of Kurzweil*, 171 A.I. 1161, 1162–63 (2007).

4. *See infra* text accompanying notes 66–68.

5. *See, e.g.*, Greg Bensinger, *ChatGPT Launches Boom in AI-Written E-Books on Amazon*, REUTERS (Feb. 21, 2023), https://www.reuters.com/technology/chatgpt-launches-boom-ai-written-e-books-amazon-2023-02-21 [https://perma.cc/AE3A-T4DZ]; Will Oremus, *He Wrote a Book on a Rare Subject. Then a ChatGPT Replica Appeared on Amazon*, WASH. POST (May 5, 2023, 2:06 PM), https://www.washingtonpost.com/technology/2023/05/05/ai-spam-websites-books-chatgpt [https://perma.cc/276Y-QLXR]; Ian Tucker, *AI Journalism is Getting Harder to Tell from the Old-Fashioned, Human-Generated Kind*, THE GUARDIAN (Apr. 30, 2023, 6:00 AM), https://www.theguardian.com/commentisfree/2023/apr/30/ai-journalism-is-getting-harder-to-tell-from-the-old-fashioned-human-generated-kind [https://perma.cc/7CZD-F2HS]. Of course, not all are impressed by ChatGPT's ability especially in the realm of poetry. *See, e.g.*, Mark Savage, *Nick Cave Says ChatGPT's AI Attempt to Write Nick Cave Lyrics "Sucks"*, BBC NEWS (Jan. 17, 2023), https://www.bbc.com/news/entertainment-arts-64302944 [https://perma.cc/855H-8N3C]; Walt Hunter, *What Poets Know That ChatGPT Doesn't*, ATLANTIC (Feb. 13, 2023, 10:10 AM), https://www.theatlantic.com/books/archive/2023/02/chatgpt-ai-technology-writing-poetry/673035/ [https://perma.cc/EAG7-PGNW].

6. *See, e.g.*, 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/3EFU-35L9]; Pranshu Verma, *AI Can Make Movies, Edit Actors, Fake Voices. Hollywood Isn't Ready*, WASH. POST (Apr. 14, 2023, 7:00 AM), https://www.washingtonpost.com/technology/2023/04/14/ai-hollywood-filmmaking-dalle [https://perma.cc/A253-WBVP]; Amanda Hoover, *AI-Generated Music Is About to Flood Streaming Platforms*, WIRED (Apr. 17, 2023, 7:00 AM), https://www.wired.com/story/ai-generated-music-streaming-services-copyright [https://perma.cc/G4KV-UQLG].

assortment of copyright infringement allegations against actors involved in the supply chain of GenAI systems.[7]

The central thesis of this Article is twofold. First, that many of the ostensible copyright concerns raised by machine production are premised on an assumption — shared widely across the field — that is simply wrong. Once the correct assumption is installed in its place, many of these problems and the ambitious infringement arguments that result from them dissipate; as it turns out, machine production raises few new policy concerns or conceptual difficulties *for copyright*. Yet, none of this is to say that machine production does not raise significant concerns for cultural policy — only that, and this is the second key claim of the Article, copyright is not the proper vehicle for addressing them.

A central premise underlying many of the copyright challenges to GenAI activities and outputs is simply misguided. Copyright orthodoxy assumes that the infringement arguments being asserted, including some of the most far-reaching ones, properly lie within the *domain* of copyright. This position accepts that the relevant GenAI-related activities are of the kind that involves copyrightable subject matter, and then

---

7. As of the time of submitting this article, there are twenty-one active lawsuits on the subject. Some of the more significant cases are: Andersen v. Stability AI Ltd. 700 F. Supp. 3d 853 (N.D. Cal. 2023) (regarding a class action brought by artists against makers of AI image generators); Complaint & Demand for Jury Trial, Getty Images (US), Inc. v. Stability AI, Inc., No. 23-cv-00135, (D. Del. Feb. 3, 2023) (alleging copyright infringement of millions of images during the AI training process); Class Action Complaint & Demand for Jury Trial, Does v. GitHub, Inc., No. 22-cv-06823 (N.D. Cal. Nov. 3, 2022) (bringing a class action against makers and distributors of an AI system for producing computer code); Complaint & Demand for Jury Trial, N.Y. Times Co. v. Microsoft Corp., No. 23-cv-11195 (S.D.N.Y. Dec. 27, 2023) (alleging improper use of millions of copyrighted news articles in the training of an AI model). For a running list of GenAI-related copyright cases, see *Lawsuits v. AI The Trial of AI: Master List of Lawsuits v. AI, ChatGPT, OpenAI, Microsoft, Meta, Midjourney & Other AI Cos.*, CHAT GPT IS EATING THE WORLD (Aug. 30, 2024), https://chatgptiseatingtheworld.com/2023/12/27/master-list-of-lawsuits-v-ai-chatgpt-openai-microsoft-meta-midjourney-other-ai-cos/ [https://perma.cc/GCF8-NTEE]. For media discussion, see, for example, Blake Brittain, *Lawsuits Accuse AI Content Creators of Misusing Copyrighted Work*, REUTERS (Jan. 17, 2023, 3:05 PM), https://www.reuters.com/legal/transactional/lawsuits-accuse-ai-content-creators-misusing-copyrighted-work-2023-01-17 [https://perma.cc/UBU2-6MAV]; Molly Enking, *Is Popular A.I. Photo App Lensa Stealing from Artists?*, SMITHSONIAN MAG., (Dec. 14, 2022), https://www.smithsonianmag.com/smart-news/is-popular-photo-app-lensas-ai-stealing-from-artists-180981281 [https://perma.cc/F4MG-9YN9]; Blake Brittain, *Getty Images Lawsuit Says Stability AI Misused Photos to Train AI*, REUTERS (Feb. 6, 2023, 12:32 PM) https://www.reuters.com/legal/getty-images-lawsuit-says-stability-ai-misused-photos-train-ai-2023-02-06 [https://perma.cc/H2KL-FVPQ]; Christopher Mims, *AI Tech Enables Industrial-Scale Intellectual-Property Theft, Say Critics*, WALL ST. J. (Feb. 4, 2023 12:00 AM), https://www.wsj.com/articles/ai-chatgpt-dall-e-microsoft-rutkowski-github-artificial-intelligence-11675466857 [https://perma.cc/GF7R-UY87]; Christopher Mims, *Chatbots Are Digesting the Internet. The Internet Wants to Get Paid*, WALL ST. J. (Apr. 29, 2023, 12:00 AM), https://www.wsj.com/articles/chatgpt-ai-artificial-intelligence-openai-personal-writing-5328339a [https://perma.cc/RLG7-YMYR]; Jonathan Stempel, *N.Y. Times Sues OpenAI, Microsoft for Infringing Copyrighted Works*, REUTERS (Dec. 27, 2023, 6:50 PM), https://www.reuters.com/legal/transactional/ny-times-sues-openai-microsoft-infringing-copyrighted-work-2023-12-27/ [https://perma.cc/7FE4-CXFN].

proceeds to analyze the claims in terms of the *scope* of protection. Copyrightable subject matter forms copyright's foundations. These subject matter rules embody the principle of what copyright is about, namely: property rights in *expression,* as opposed to other kinds of informational or tangible objects and activities.[8] By contrast, scope doctrines, such as the infringement test and the fair use defense, form the next floor in copyright's conceptual structure. If and only if a specific activity or object lies within copyright's subject matter domain, does the next inquiry arise of whether it is of the kind that falls within the scope of the owner's right to exclude.[9] Mainstream analysis of GenAI infringement rushes too quickly past the foundational subject matter inquiry and focuses its entire energy on scope debates.

This Article argues that domain or subject matter principles have a central role to play in GenAI copyright cases. Some of the broadest and most practically impactful infringement arguments should fail at the domain threshold, never reaching the stage of scope inquiries. Ignoring subject matter rules fails to take seriously what the intangible objects of copyright protection are, namely: forms of expression. The result is that cases that should be decided on clearcut threshold principles sink in a quagmire of complex doctrinal debates, such as those involving the notorious nuances of the fair use doctrine. The failure to track copyright's basic structural principles leads to unnecessary confusion, undue complexity, and doctrines increasingly harder to administer and fraught with greater probability of error.

The GenAI copyright conflicts generate a litany of unorthodox infringement arguments. Unorthodox infringement arguments try to establish broad copyright liability based on various activities related to the production and deployment of GenAI systems. What makes these arguments unorthodox is their attempt to stretch copyright liability, in different ways, beyond the core case in which protected expression is potentially exposed to human consumption through one of the activities included in the owner's exclusive rights.[10] There are two sources for unorthodox GenAI infringement arguments; subject matter principles have a crucial role to play with respect to each. First, broad infringement arguments are driven by concerns internal to copyright. Copyright owners regard their interests as endangered by GenAI expressive production and hope to receive a share of the considerable value created by it. Alongside conventional infringement arguments, these owners make newer, broader claims as means for reaching the deeper pockets and stronger control of actors who are located at central junctures of

8. See 17 U.S.C. § 102(a).
9. See *infra* text accompanying notes 92–102.
10. *See* 17 U.S.C. § 106.

producing GenAI systems.[11] The role of subject matter rules here is to safeguard copyright's internal policy balance: to erect a firm barrier against extending the right to exclude to material and activities where it does not belong. Second, ambitious infringement arguments are fueled by and garner sympathy due to broader cultural policy concerns that are external to copyright's focus. Anticipating various policy problems that will arise as a result of GenAI gaining dominance in markets for expressive production, observers hope to wield copyright as a weapon for addressing these problems.[12] The role of subject matter principles here is to prevent the use of copyright for addressing external policy concerns, however genuine, that are beyond its ken. Under this more external function, subject matter rules keep copyright focused on problems close to its core concerns and for which it possesses adequate institutional tools.

Consider first the internal dynamics of copyright. Initially, the main copyright questions triggered by the rise of GenAI were primarily about authorship and rights.[13] This is hardly surprising, given that modern copyright, since its inception, has been centrally preoccupied with the imagery and ideology of individual authorship.[14] The new image of machine "authors" and their, or their human operators', potential "machine-author-rights" was the immediate focus of fascination, even if the

---

11. Oren Bracha, *Generating Derivatives: AI and Copyright's Most Troublesome Right*, 25 N.C. J.L. & TECH. 345, 355 (2024).

12. A good example of broad infringement arguments motivated by general cultural policy concerns is the argument that GenAI infringes copyright by appropriating the "style" of a specific artist. On close examination, this argument is revealed to be motivated by fears of cost-effective GenAI displacing human creators in markets for new works, rather than copyright's concern about an artist's ability to recoup the cost of creating a specific work. *See infra* text accompanying notes 213–15.

13. *See generally* Karl F. Milde Jr., *Can a Computer Be an "Author" or an "Inventor"?*, 51 J. PAT. OFF. SOC'Y 378 (1969); Pamela Samuelson, *Allocating Ownership Rights in Computer-Generated Works*, 47 U. PITT. L. REV. 1208 (1986); Annemarie Bridy, *Coding Creativity: Copyright and the Artificially Intelligent Author*, 5 STAN. TECH. L. REV. 1, 26 (2012); Annemarie Bridy, *The Evolution of Authorship: Work Made by Code*, 39 COLUM. J.L. & ARTS 395 (2016); Robert C. Denicola, *Ex Machina: Copyright Protection for Computer-Generated Works*, 69 RUTGERS L. REV. 251 (2016); 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 (2016); Shlomit Yanisky-Ravid, *Generating Rembrandt: Artificial Intelligence, Copyright, and Accountability in the 3A Era — The Human-Like Authors are Already Here — A New Model*, MICH. ST. L. REV. 659 (2017); Daniel Gervais, *The Machine as Author*, 105 IOWA L. REV. 2053 (2020); Patrick Goold, *Artificial Authors: Case Studies of Copyright in Works of Machine Learning*, 67 J. COPYRIGHT SOC'Y U.S.A. 427 (2020); Vicenc Feliu, *Our Brains Beguil'd: Copyright Protection for AI-Created Works*, 25 INTELL. PROP. & TECH. L. J. 105 (2021); P. Bernt Hugenholtz & Joao Pedro Quintais, *Copyright and Artificial Creation: Does EU Copyright Law Protect AI-Assisted Output?*, 52 INT'L REV. INTELL. PROP. & COMPETITION L. 1190 (2021).

14. *See, e.g.*, Martha Woodmansee, *The Genius and the Copyright: Economic and Legal Conditions of the Emergence of the "Author"*, 17 EIGHTEENTH-CENTURY STUD. 425, 429 (1984); MARK ROSE, AUTHORS AND OWNERS: THE INVENTION OF COPYRIGHT 135 (1993); Oren Bracha, *The Ideology of Authorship Revisited: Authors, Markets, and Liberal Values in Early American Copyright*, 118 YALE L.J. 186 (2008).

relevant machines did not take quite the human form as the authorial holographic Doctor in Star Trek Voyager.[15] However, it quickly became clear that GenAI is no more a "romantic author" than humans are.[16] Like people, GenAI does not create ex nihilo.[17] To create, both people and machines have to stand on the shoulders of giants: to use and learn from the expressive works of many who came before them. While common to all creation, the cumulative character of creation is more viscerally visible with machine creation technologies, at least those that we have now. GenAI systems involve an indispensable stage of training or learning in which machines must be exposed to large quantities of existing expressive works. GenAI can only generate after it "sees" immense amounts of works and extracts from them common patterns.[18] In this sense, machine generation takes all the mysticism out of the process of creation. Ex nihilo fantasies are hardly possible to entertain when a process of training by exposure to prior works discretely and systematically precedes the stage of creation. The giants (of all sizes), so to speak, are sitting in the room.

The visibly cumulative character of machine creation, together with its emergent source of social value, has given rise to a flood of disputes and claims. Artists are concerned about the use of their works to train machines that they fear might outcompete them.[19] Getty Images is furious about unlicensed use of copyrighted images in its portfolio for training purposes.[20] And the New York Times wants the power to control the training of GenAI systems on its copyrighted content that may sometimes result in generated output similar to that content.[21] In these lawsuits various distinct arguments for copyright infringement swirl together. Some are hardly novel. There seems to be little doubt that the generated output of a GenAI system that is "substantially similar" to a copyrighted work may infringe copyright.[22] Other

15. *See Author, Author (Star Trek: Voyager)*, WIKIPEDIA, https://en.wikipedia.org/wiki/Author,_Author_(Star_Trek:_Voyager) [https://perma.cc/E9KD-RUY2].

16. The "romantic author" is an ideological construct of creation as an individualist act, imagined as an atomistic individual producing something completely new out of her/his mind. The ideology of romantic authorship has had a complex relationship with modern copyright since the inception of the field in the early eighteenth century. *See* Bracha, *supra* note 14, at 200.

17. *See infra* text accompanying notes 55–63(describing the training and generation stages of producing a GenAI system).

18. *See infra* text accompanying notes 57–58.

19. *See* Andersen v. Stability AI Ltd., 700 F. Supp. 853, 853 (N.D. Cal. 2023).

20. *See* Complaint & Demand for Jury Trial at 1, Getty Images (US), Inc. v. Stability AI, Inc., No. 23-cv-00135 (D. Del. Feb. 3, 2023).

21. *See* Complaint & Demand for Jury Trial at 3, N.Y. Times Co. v. Microsoft Corp., No. 23-cv-11195 (S.D.N.Y. Dec. 27, 2023).

22. *See* Arnstein v. Porter, 154 F.2d 464, 473 (2d Cir. 1946). The main challenge with respect to such cases is ascertaining who the agent directly responsible for the infringement is and which other entities in the AI production supply chain may have derivative liability.

infringement arguments are much more ambitious: by asserting broad claims of liability, they stretch copyright law's frontiers and challenge its basic assumptions.

Arguments of the latter kind come in two forms: upstream and downstream. The process by which GenAI generates expression is best understood as a supply chain involving discrete stages and various actors.[23] Upstream arguments focus on the earlier stages of this process revolving around the training of a model. The most common claim in this set is that creating training copies — the reproduction of digital files representing copyrighted works during the GenAI training process — is itself copyright infringement.[24] Downstream arguments target the later stages of the process by attempting to expand the liability that applies to generated expressive materials even when these bear only remote or diffused similarity to existing copyrighted works. A representative, and widespread, version of this argument is that generated output infringes copyright by appropriating a creator's recognizable expressive "style."[25]

How should these more ambitious, and sometimes novel, claims for GenAI copyright infringement be evaluated? Two common, yet opposing, responses should be avoided: first, mechanical or blind, applications of existing rules and precedents without any examination of how to serve their underlying purposes in the new context, and second, attempts to rewrite copyright law from scratch in the face of the new challenges posed by machine production. In place of either extreme, we need instead to turn to the basic principles of copyright, ask what purposes those principles are designed to serve, and how best, in light of these purposes, to apply existing rules to the new context. Examining the emergent phenomena in light of the field's fundamental principles generates clear answers to the new challenges. Many of the ambitious infringement claims, on both the upstream and downstream sides, run against copyright's foundational subject matter principles and would undermine their fundamental purposes, and thus, should be rejected.

The copyright principle most crucial to resolving GenAI infringement challenges is what we may call the "spillovers principle."[26] The spillovers principle is a fundamental structural feature that has been modern copyright's bulwark against the specter of entangling social

---

See Michael Goodyear, *Infringing Information Architectures*, 58 U.C. DAVIS L. REV. (forthcoming 2025) (manuscript at 45), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4747940 [https://perma.cc/T8XS-G7KY].

23. *See* Katherine Lee, Feder A. Cooper & James Grimmelmann, *Talkin' 'Bout AI Generation: Copyright and the Generative-AI Supply Chain*, J. COPYRIGHT SOC'Y U.S.A. (forthcoming 2024) (manuscript at 7) (on file with author).

24. *See infra* text accompanying notes 122–124.

25. *See infra* text accompanying notes 192–195.

26. *See* Brett M. Frischmann & Mark A. Lemley, *Spillovers*, 107 COLUM. L. REV. 257, 258 (2007) (coining and defining the term "spillover").

knowledge in the manacles of private property since its inception.[27] Under the spillovers principle, the right to exclude conferred by copyright is strictly limited in both *domain* and *scope*.[28] Copyright has never been a plenary right to exclude from all valuable aspects or uses of a work. And this restriction is a feature, not a bug. Copyright is based on the assumption that full internalization of the work's value through expansive rights to exclude from all its valuable uses is not a desirable goal.[29] Under the domain side of the spillovers principle, copyright is strictly limited to one aspect of an information good: expressive forms. By design, everything else that is often bundled alongside expression in copyrighted works — be it information or knowledge, functional elements, material aspects, and even certain structural expressive features — is to be spilled over into the public domain.[30] This is the case no matter how valuable the non-protectable elements are, indeed even if they are more valuable than the expression.

When viewed through the lens of the spillovers principle, the entire GenAI legal and policy debate at the upstream level is deeply misconceived. Both critics and defenders of the reproduction involved in training copies take for granted that such copies constitute *prima facie* infringement. The reason? A copy is a copy. Since training copies embody exact reproduction of the physical patterns that represent the informational content of works, they must be infringing.[31] The debate then focuses on the question of whether the reproduction is, nevertheless, to be excused under the fair use defense.[32]

This shared ground of this debate is categorically wrong. Under copyright's subject matter rules, the proper domain of the field is strictly limited to expressive forms. Copyright is about the production and consumption of the value of expression qua expression. The GenAI training process is equivalent to a type of learning that has always been permitted under modern copyright: a process of extraction of meta-information from expressive works that then enables the production of new and different expression.[33] The only difference is that machine learning incidentally involves physical reproduction. It requires the making of a physical object, an object from which no human would ever access the expressive content of the work, necessary to extract the meta-information.[34] Focusing on this difference to label an otherwise allowed activity of learning as infringing is succumbing to a fallacy of physicalism. It assumes that copyright cares about physical facts as

27. *See infra* text accompanying notes 82–84.
28. *See infra* text accompanying notes 92–102.
29. *See infra* text accompanying notes 103–112.
30. *See* 17 U.S.C. § 102(b).
31. *See infra* text accompanying notes 121–122.
32. *See infra* text accompanying note 124.
33. *See infra* Section IV.A.
34. *See infra* text accompanying notes 55–61 (describing the training process).

such, rather than access to, and use of, expressive value. Copyright's domain, however, is expressive value, not physical objects. Creating a physical object from which no human will ever access the expressive value of the work simply does not involve any copyrightable subject matter. It is the equivalent of using a book as a doorstop. Notwithstanding the physicalist fact of reproduction, training copies involve no reproduction of copyrightable subject matter and therefore cannot infringe.[35] This is not owing to scope-type considerations at the backend, such as fair use. Non-expressive training copies simply do not infringe from the outset, due to the most basic first principles of copyright that determine what subject matter lies within its domain in the first place.

The spillovers principle similarly dismisses expansive arguments at the downstream level; specifically, those asserting that the reproduction of the "style" of a particular creator, rather than any specific expressive work, is infringing. Arguments concerning the appropriation of style target, once more, subject matter that lies outside copyright's domain. This is so for two interlocking reasons. First, "style" is a made-up information good, fabricated by combining elements, conceived at a highly abstract level, taken from different expressive works. But copyright applies to specific works rather than a corpus of works.[36] It does not recognize such a cross-work informational object. Second, arguments about style, exactly because they cannot establish similarity to the concrete expressive patterns of any specific work, rely on the appropriation of highly abstract elements and patterns. But these informational elements are ones that copyright's subject matter rules designate as "ideas" and place outside the field's domain.[37]

Turning to broader cultural policy concerns, another function of subject matter rules is revealed. One might ask: why are such unorthodox infringement arguments being asserted against GenAI? Why not simply stick with traditional infringement claims, pertaining to producing or disseminating substantially similar, expressive works in forms accessible to humans? Part of the answer is that such arguments are driven by a desire to use copyright to address the fundamental policy concerns and anxieties that arise in the wake of the socially disruptive effect of AI. In the cultural realm, these concerns mainly take the form of fears of GenAI displacing human authors from markets for expression that may result in three unfortunate effects: the dissipation of sources of income in creative industries, the diminishment of opportunities for accessing the inherent value of expressive activities, and the weakening of sources for paradigm-breaking creative innovation.[38]

---

35. *See infra* Section IV.D.
36. *See infra* text accompanying notes 206–207.
37. *See* 17 U.S.C. § 102(b).
38. *See infra* text accompanying notes 220–231.

Broad copyright claims are used as an attempt to stop or at least slow down these disconcerting prospects.[39] The general policy concerns surrounding the rise of GenAI in markets for expression may be genuinely important. Copyright, however, is the wrong legal field for addressing them. Copyright was designed as a remedy for a specific information-policy problem and was endowed with specific institutional tools for alleviating it. These tools are ill-suited for addressing other, very different social policy concerns for which they were not designed. Subject matter rules play a more external function here. Limiting copyright's application only to cases where its relevant subject matter is implicated ensures that the field governs only the kind of policy problems it was designed and equipped to handle.

This article proceeds in five parts. Part II supplies the necessary background on the technical operation and institutional context of GenAI. Part III explains the spillovers principle as a deep structural feature of modern copyright and its embodiment in domain subject matter rules. The following two Parts then apply copyright's subject matter rules to the twin categories of unorthodox infringement claims. Part IV analyzes upstream infringement arguments about reproduction in training copies. It explains why such arguments should be dismissed at the front gate of subject matter rules, rather than through the backdoor of fair use, and how this analysis behooves us to reconsider, more generally, existing case law on non-expressive uses of copyrighted works. Part V explains why downstream arguments about appropriation of style similarly fail on subject-matter grounds. Part VI then zooms out. It discusses some of the more fundamental policy concerns that arise in view of the specter of machine generation replacing a significant share of market-backed human creativity, and copyright's inadequacy for addressing those concerns. Part VII concludes.

II. THE IMITATION GAME: GENERATIVE AI AND EXPRESSIVE GOODS

To analyze the copyright law and policy of GenAI, one must first understand how this technology generates expressive works. This Part provides a simplified version of the key features of GenAI as it operates in the field of creating expressive goods.

GenAI can be explained by locating it within its more general technological field and contrasting it with adjacent subfields. GenAI is a subfield of Machine Learning which is itself a subfield of the broader

39. Of the two sets of broad infringement arguments, those targeting the upstream production stages are the more potent: they strike the GenAI production process at its root and they apply to all systems, irrespective of whether they generate expressive output at all. However, downstream arguments too represent attempts to dramatically expand copyright beyond its traditional boundaries, one that if successful is likely to bleed beyond the GenAI context.

area of Artificial Intelligence. Artificial Intelligence is commonly defined as the field of developing machines (today this primarily means digital computers) that exhibit intelligence by mimicking the problem-solving and decision-making abilities distinctive of the human mind.[40] Computerized machines playing chess, processing natural language, and making decisions related to driving a car are a few examples. There are various approaches to designing AI. An expert system approach, that was popular in earlier phases, is based on processing information by executing complex systems of pre-given rules or decision trees.[41] Machine learning is a competing approach that has proven tremendously fruitful in the recent few decades.[42] The distinctive feature of this approach is its learning component. Unlike expert systems, machine learning systems do not simply follow a set of pre-given rules, no matter how complex. Instead, such systems learn — that is to say, they apply algorithms to sets of relevant sample data in order to build their own parameters for making the desirable decisions or predictions.[43]

A rapidly growing subset within machine learning is GenAI whose distinctive feature is generating new information goods.[44] Some AI systems do not generate new information goods, or at least not of the kind that is consumable by humans. Chess-playing, autonomous cars, and face recognition systems are some examples. The main purpose of GenAI, in contrast, is to generate new information goods; anything from musical compositions to price predictions. The distinction is hazy,

40. *Artificial Intelligence*, IBM, https://www.ibm.com/design/ai/basics/ai/ [https://perma.cc/TES8-9GJP] (defining AI as "[a]ny system capable of simulating human intelligence and thought processes"); *see also* STUART RUSSELL & PETER NORVIG, ARTIFICIAL INTELLIGENCE: A MODERN APPROACH 1 (4th ed. 2022) (emphasizing the field of AI is concerned with "*building* intelligent entities — machines that can compute how to act effectively and safely in a wide variety of novel situations"). As machines become increasingly powerful and exhibit abilities far surpassing those of humans in many fields, the element of mimicking the human mind is sometimes dropped from the definition. This leaves exhibiting intelligence as the key element of the concept and sharpens already difficult questions about what exactly intelligence is. *See* RUSSELL & NORVIG, *supra*, at 19–22 (discussing competing concepts of intelligence as fidelity to human performance or as more general rationality).

41. *See* RUSSELL & NORVIG, *supra* note 40, at 42.

42. *See* JUGAL KALITA, MACHINE LEARNING IN THEORY AND PRACTICE 2–4 (2022) (explaining the concept of machine learning).

43. ETIENNE BERNARD, INTRODUCTION TO MACHINE LEARNING 1 (2021).

44. MOHAK AGARWAL, GENERATIVE AI FOR ENTREPRENEURS IN A HURRY ch. 1 (2023) ("While traditional AI is designed to recognize or classify existing data, generative AI is able to generate novel and diverse outputs based on a given set of input parameters or conditions."); BERNARD, *supra* note 43, at 14 (describing generative modeling in AI as "the most difficult unsupervised learning task" because it requires models "to learn how to generate examples that are similar to the training data").

but it captures an important and increasingly impactful subset of AI applications.[45]

Another hazy but useful distinction is the one between GenAI and more traditional data-mining technology. Data-mining is primarily focused on indexing and extracting useful meta-information out of data sets.[46] Internet search engines (both general and niche) and Google Books, which allows searching the texts of physical books, are examples of data-mining-based systems.[47] GenAI's training stage is in fact a form of data-mining. The distinctive feature of GenAI, however, is that it uses the mined data to create new information goods rather than to primarily index, analyze, search or even retrieve preexisting information or patterns. Accordingly, the training stage is followed by a generation stage.

Finally, within GenAI, some systems are distinct in that they are designed to generate new *expressive goods* in various media. Except in a trivial and incidental manner, a GenAI system whose main purpose is to generate price-predictions or new technological inventions is not an expressive-goods-generating system. But an image-, text- or music-generating system is. The product of such GenAI systems is often referred to as "generative art."[48] For current purposes, it is better to use the term expressive goods.[49] The focus of this Article is expressive goods GenAI.[50]

How does GenAI work?[51] As explained by Lee, Cooper & Grimmelmann, the production and operation of such systems is composed of

---

45. One reason why the distinction is hazy is that meta-information is information. Meta-information is information about information. And almost any AI system generates some new meta-information. Perhaps one way of sharpening the GenAI concept is to reformulate it as encompassing systems whose main purpose is to generate new information comparable to that in their training set. See BERNARD, *supra* note 43, at 14.

46. *See* JIAWEI HAN, MICHELINE KAMBER & JIAN PEI, DATA MINING: CONCEPTS AND TECHNIQUES 2 (4th ed. 2023) (defining data mining as "the process of discovering interesting patterns, models and other kinds of knowledge in large data sets"); *see also* Matthew Sag, *The New Legal Landscape for Text Mining and Machine Learning*, 66 J. COPYRIGHT SOC'Y U.S.A. 291, 294–301 (2019).

47. HAN ET AL., *supra* note 46, at 17.

48. *See* Margaret A. Boden & Ernest A. Edmonds, *What is Generative Art?*, 20 DIGIT. CREATIVITY 21, 29–30 (2009) (defining the term "generative art" as applying to cases where "the artwork is generated, at least in part, by some process that is not under the artist's direct control").

49. Modern copyright law applies to expression and formally avoids thresholds of aesthetic merit or being a work of art. *See* Bleistein v. Donaldson Lithographing Co., 188 U.S. 239, 251 (1903).

50. To avoid cumbersome language from this point onward I will be using "GenAI" to refer to GenAI that generates expressive goods, unless I say otherwise.

51. For an illuminating and accessible resource on this subject, see Stephen Wolfram, *What Is ChatGPT Doing . . . and Why Does It Work?*, STEPHEN WOLFRAM WRITINGS (Feb. 14, 2023), https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work [https://perma.cc/T4NK-C94R]. *See also* Pamela Samuelson, *Generative AI Meets Copyright*, 381 SCIENCE 158, 159 (July 2023) (describing the process of training a GenAI model).

a complex supply-chain.[52] The process can be conceptually broken down into multiple stages.[53] There are likely to be different actors at play in each of these stages and the institutional models may differ greatly.[54] A full examination of the copyright implications of the entire GenAI supply chain would require separate, context-specific analysis of the activities of each of these actors in each of the links of the supply-chain. Such full examination is beyond the scope of this Article. The focus here is on the two central infringement arguments that have been raised with respect to either end of the supply-chain: upstream training copies and downstream copying of style. Consequently, for current purposes, we can simplify by tentatively reducing the complex GenAI supply chain into two stages: upstream training and downstream generation.

To simplify, assume a text-based-system, such as ChatGPT.[55] Such a system is text-based in three ways: its training set, user prompts, and generated output. The central upstream component in the production of such a system is the training of a Large Language Model ("LLM").[56] In the training stage, the model is created by extracting meta-information out of a large training set consisting of various texts. Generally, the larger and more inclusive the training set, the better the results.[57] One may say that the system "reads" the texts. In more technical terms, however, the process is as follows. To be accessible to the system, digital files representing the texts in the training set are reproduced. The system accesses and analyzes the files. In this analysis, the texts are broken into small fundamental units called tokens.[58] The system then applies various functions and operations to the sequences of tokens to identify and extract patterns in them. The outcome of this process, one which is far more complex than the simplified version presented here, is a large array of parameters, values, or weights.[59] What these parameters

52. *See* Lee et al., *supra* note 23, at 36.

53. *Id.* at 5–6 (proposing to analyze the AI supply chain as consisting of eight different stages).

54. *Id.* at 32 (calling attention to the question of which actors are involved in each stage of the AI supply chain).

55. The term Large Language Models is often used to refer to machine learning models consisting of complex neural networks that are applied to text. However, the term is also used sometimes with respect to similar models applied to other media such as images and music. *See, e.g.*, AGARWAL, *supra* note 44, at ch. 2.

56. The training or production stage can be further conceptually divided into at least five substages as follows: data creation, data set collection and curation, model (pre-)training, model fine tuning, and model alignment. *See* Lee et al., *supra* note 23, at 36.

57. *See* BERNARD, *supra* note 43, at 38 (observing that typical ways to improve performance in machine learning is to add more data and diversify its origin).

58. One may think about these tokens as words, but in reality, the units will not map exactly onto words and may be smaller or larger. *See* Wolfram, *supra* note 51, at 45 (explaining that "ChatGPT does not deal with words, but rather with 'tokens' — convenient linguistic units").

59. One complication is that the training process may be supervised or unsupervised. *See* BERNARD, *supra* note 43, at 9–14.

represent is a complex set of probabilities that describe and allow "guesses" on which specific token is most likely to follow a given sequence of tokens.[60] This set of parameters is the output of the training process, or the model.[61] Following its training, the model is deployed by being incorporated into a system as part of a product or a service.[62]

Next comes the generation stage in which the system generates new information goods. In this stage, the system accepts inputs — commonly referred to as prompts — from users and generates corresponding outputs. A user's prompt, in our basic scenario, is itself a text string, say "write me a short story about alienation in modernity." After the prompt is submitted, the system attempts to infer the "correct" response to the user's prompt.[63] Inference is a probabilistic process of composing a sequence out of tokens in response to a prompt. The prompt string is simply treated as an initial sequence of tokens. The system then applies its set of parameters to guess the most probable next token and repeats the process until it completes the sequence. The output of this inference process is a new text string. In our basic scenario, this new text, that the user (if the system is a good one) may recognize as a short story that fits her request, is nothing more than a series of probabilistic guesses about a sequence of tokens.

The basic principles are extendable, *mutatis muntandis*, to other media. To extend a GenAI system to generate output in media other than text, two main elements are necessary. First, the system needs a training process like the one described above with a training set composed of data of the relevant media. In principle, any digitizable media — for example, image, sound or video — could be subject to the process of tokenization (being broken down into units) and parameter extraction, where the parameters represent probabilistic sequence patterns or a model.[64] Second, the system needs some way of connecting text prompts to sequences of the relevant media: a way of performing the inference stage that starts for example with the text "evil black cat" and proceeds to constructing a pictorial sequence that users experience as a matching image. This additional element of bridging the different media of the prompt and the output is achieved by yet another layer of machine learning training.[65] In this process, the machine extracts

60. BERNARD, *supra* note 43, at 22 (describing the training of language models whose purpose is to predict the next word after a given sequence).

61. A human intervention step commonly follows this stage, imposing external constraints, for example, making sure that the system does not produce results corresponding to certain prompts.

62. *See* Lee et al., *supra* note 23, at 53.

63. BERNARD, *supra* note 43, at 11 (explaining the "*inference phase*").

64. AGARWAL, *supra* note 44, at ch. 2.

65. *See* Jorge Agnese, Jonathan Herrera, Haicheng Tao & Xingquan Zhu, *A Survey and Taxonomy of Adversarial Neural Networks for Text-to-Image Synthesis* 2 (unpublished manuscript) (on file with arXiv), https://arxiv.org/abs/1910.09399 [https://perma.cc/AMA8-QCTG] (discussing machine learning methods for text-to-image synthesis).

patterns that connect forms in one media — textual labels like "cat" or "black" — to sequences in another, for example, image sequences. The outcome of this additional layer of training is yet again a set of parameters that represent probabilistically the connections between patterns in the two media. Armed with these two sets of data produced by the training process, the system can proceed to a probabilistic inference process that constructs output sequences in the relevant media as a response to textual prompts. In principle, similar design elements could apply to connect any prompt and output media. With proper training, a GenAI system could, for example, take image or sound input as prompts and generate text or video output.[66]

The result of all of this is a staggering potential for machine generation of expressive goods in a variety of media and in response to various inputs.[67] What we are beginning to experience is a process in which this potential explodes into realization. The possible applications are many and some, such as "deep fakes," have little to do with markets for expressive goods.[68] Our focus here, however, is on GenAI expressive goods as used in markets for expression, markets in which people pay for and gain access to expressive goods in order to consume and enjoy their expressive value.

In two respects, we are at a tipping point in the impact and significance of GenAI in markets for expressive goods. First and foremost, the power of GenAI systems and the quality of their output is reaching the point where their expressive products can serve as adequate substitutes for human-created goods,[69] a process which is certain to expand in coverage and intensity. Further, rather than a binary division of machine and human-created goods, there are signs of development of hybrid models. By using GenAI for various segments of an expressive project while also combining its output with human contribution,

66. Models capable of processing information from multiple types of data are referred to as "multi-modal models," for example, ChatGPT 4 and Google's Gemini. *See* Cole Syrykr, *What Is Multimodal AI?*, IBM (July 15, 2024), https://www.ibm.com/think/topics/multimodal-ai [https://perma.cc/8WYK-9HA2].

67. I am using the term "expressive" in a narrow technical sense. The term denotes that GenAI output can function as good substitutes for enjoyment by humans of the value of other expressive works, like images, videos, texts or music. The use of the term does not imply any claim that the relevant information goods are expressive in the sense that they involve a process of creation or a creating agent which are equivalent to those in the case of producing human expression.

68. *See, e.g.*, Shannon Bond, *AI-Generated Deepfakes are Moving Fast. Policymakers Can't Keep Up*, NPR (Apr. 23, 2003), https://www.npr.org/2023/04/27/1172387911/how-can-people-spot-fake-images-created-by-artificial-intelligence [https://perma.cc/M89F-YHEL]; *see also* Bobby Chesney & Danielle Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CALIF. L. REV. 1753, 1756–59 (2019).

69. BERTIN MARTINS, ECONOMIC ARGUMENTS IN FAVOUR OF REDUCING COPYRIGHT PROTECTION FOR GENERATIVE AI INPUTS AND OUTPUTS 16 (2024), https://www.bruegel.org/system/files/2024-04/WP%2009%20040424%20Copyright%20final_0.pdf [https://perma.cc/C7C5-47X2] ("GenAI reduces media production costs and triggers price, quantity and substitution effects.").

creators can now cut dramatically on production cost and time.[70] Second, various business models for exploiting GenAI expressive goods are appearing. Such models are not restricted to the straightforward sale or licensing of copies of GenAI-generated or hybrid expressive goods. Instead, they involve other configurations that are both more decentralized in some respects and more centralized in others. Such models include syndication of the services of expressive GenAI engines to downstream services and end-users. A dominant example is the proliferation of image generating applications or services that allow end-users to obtain GenAI-generated images to their specifications.[71] Another important model is integration of GenAI expression abilities into existing services and tools, such as internet search engines or graphics software.[72]

The predictable upshot of the growing penetration of GenAI into markets for expressive goods is an intense disruption of the way these markets operate.[73] Since the market is the predominant institutional form through which our society organizes the production, access, and

---

70. *See, e.g.*, Greg Bensinger, *ChatGPT Launches Boom in AI-Written E-Books on Amazon*, REUTERS (Feb. 21, 2023, 3:43 PM), https://www.reuters.com/technology/chatgpt-launches-boom-ai-written-e-books-amazon-2023-02-21/ [https://perma.cc/8FVL-YDE6]; Travis Diehl, *Mimicking the 19th Century in the Age of A.I.*, N.Y. TIMES (May 3, 2023), https://www.nytimes.com/2023/05/03/arts/design/ai-makes-nostalgic-images.html [https://perma.cc/H8G3-QTDW].

71. *See, e.g.*, DALL-E 2, https://openai.com/index/dall-e-2/ [https://perma.cc/35K9-M4N6]; MIDJOURNEY, https://www.midjourney.com/home [https://perma.cc/P6YN-QUKF]; ARTBREEDER, https://www.artbreeder.com/ [https://perma.cc/A99V-YBE5]; *see also* Arianna Johnson, *Here Are The Best AI Image Generators*, FORBES (Apr. 28, 2023, 5:37 PM), https://www.forbes.com/sites/ariannajohnson/2023/04/28/here-are-the-best-ai-image-generators/ [https://perma.cc/V5C5-XLLS].

72. *See, e.g.*, Jeffrey Dastin, *Microsoft Packs Bing Search Engine*, *Edge Browser with AI in Big Challenge to Google*, REUTERS (Feb. 7, 2023, 8:28 PM), https://www.reuters.com/technology/microsoft-infuse-software-with-more-ai-google-rivalry-heats-up-2023-02-07 [https://perma.cc/Q3X5-CNK3]; Katherine Hamilton, *Amazon Launches AI Platform Aimed At Corporate Customers — Joining Google and Microsoft in AI Race*, FORBES (Apr. 13, 2023, 11:04 AM), https://www.forbes.com/sites/katherinehamilton/2023/04/13/amazon-launches-ai-platform-aimed-at-corporate-customers-joining-google-and-microsoft-in-ai-race/ [https://perma.cc/Q3X5-CNK3]; Nico Grant, *Google Builds on Tech's Latest Craze with Its Own A.I. Products*, N.Y. TIMES (May 10, 2023), https://www.nytimes.com/2023/05/10/technology/google-ai-products.html [https://perma.cc/XN9R-F3N8]; Oliver Darcy, *News Publishers Sound Alarm on Google's New AI-infused Search, Warn of 'Catastrophic' Impacts*, CNN (May 15, 2024, 7:04 AM), https://www.cnn.com/2024/05/15/media/google-gemini-ai-search-news-outlet-impact/index.html [https://perma.cc/W69Z-A6VH].

73. *See, e.g.*, Bensinger, *supra* note 70; Jaclyn Paiser, *The Rise of the Robot Reporter*, N.Y. TIMES (Feb. 5, 2019), https://www.nytimes.com/2019/02/05/business/media/artificial-intelligence-journalism-robots.html [https://perma.cc/78HE-22QS]; 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/3EFU-35L9]; Amanda Hoover, *AI-Generated Music Is About to Flood Streaming Platforms*, WIRED (Apr. 17, 2023, 7:00 AM), https://www.wired.com/story/ai-generated-music-streaming-services-copyright/ [https://perma.cc/G4KV-UQLG]; Darcy, *supra* note 72.

use of expressive goods, these human activities will be deeply shaped by the transformation. And since copyright law is the most important institutional mechanism that connects the production and use of expressive works to markets, many of the struggles and dilemmas that arise are being laid at its doorstep.

III. THE SPILLOVERS PRINCIPLE

The two most important unorthodox arguments of copyright infringement directed against GenAI target the two distinct stages of its operation. The first aims at the upstream stage of the GenAI supply chain by arguing that unauthorized digital reproduction of copyrighted works during the training process is itself copyright infringement.[74] The second trains its sights on downstream generation by asserting that certain GenAI output, while not resembling any concrete work, nevertheless infringes copyright by recognizably copying the "style" of certain creators.[75] Although the arguments are very different, both implicate and ultimately are resolved by a fundamental principle that has been constitutive of modern copyright since its inception: the spillovers principle. This Part explains the spillovers principle. The two following Parts apply the principle to the questions of infringement by training copies and by appropriating style.

One of copyright's most fundamental tenets is the spillovers principle.[76] Under this principle, copyright is strictly limited to a specific domain — that of expressive forms — and to a circumscribed scope within that domain. All elements of an expressive work that do not fall within this domain and scope are allowed, by design, to "spillover" and remain unowned and uncontrolled, free for all to use. This is true irrespective of the value and centrality of unprotected informational elements, even if the value of such elements far exceeds that of protectable ones. Newton, had he published and copyrighted his Philosophiae Naturalis Principia Mathematica a half-century later, could not have

74. *See* Complaint at 1, Getty Images v. Stability AI, Inc., 23-cv-00135 (D. Del. Feb. 3, 2023) (requesting relief for copyright infringement on the basis of the allegation that defendant trained GenAI with copyright images and as part of the process caused those images to be stored at and incorporated into its system).

75. *See* Complaint at 2, Andersen v. Stability AI Ltd., 23-cv-00201 (N.D. Cal. Aug. 12, 2024) (requesting relief for copyright infringement for "works generated by AI Image Products 'in the style' of a particular artist").

76. The term "spillovers" is borrowed from Frischmann & Lemley, *supra* note 26, at 258, and the work of Brett Frischmann more generally. *See* Brett M. Frischmann, *Speech, Spillovers, and the First Amendment*, 2008 U. CHI. L. F. 301; Brett Frischmann, *Spillovers Theory and Its Conceptual Boundaries*, 51 WM. & MARY L. REV. 801 (2009), Although the general argument here is similar to theirs, my usage of the term is somewhat different.

stopped anyone from reproducing and using the theories developed in it despite them being the main value of the work.[77]

The purpose of the spillovers principle is to allay the deep concern that copyright's beneficial goal — whether understood as supporting creation or rewarding creators — might come at too heavy a price of restricting the flow of knowledge and the cumulative development of learning and culture. Its origins date back to the earliest days of modern copyright.

Premodern copyright was founded on the purpose of restricting the circulation of knowledge.[78] Since at least as early as the 1557 Charter of the Stationers' Company, granted as a response to the dissemination of "seditious and heretical books rhymes and treatises," and for at least another century, the purpose of proto-copyright and the logic of its institutional structure were tightly wrapped with state censorship.[79] The axiomatic assumption was, in the words of a 1643 petition of the Stationers' Company to Parliament, that "the first and greatest end of order in the Presse, is the advancement of wholesome knowledge."[80] In the wake of the political crisis in England that resulted in the demise of the old regulation of the press system and the rise of modern copyright, the assumption that copyright is designed to restrict the circulation of knowledge was inverted. The 1710 Statute of Anne,[81] marking the beginning of modern copyright, was enacted after the old censorial grounding lost its traction.[82] It expressly grounded copyright, not in controlling knowledge, but in the diametrically opposed purpose of the "encouragement of learning."[83]

Modern copyright was thus born with an inherent tension built into it: it was a mechanism of private control of expression, backed by state sanction, yet was officially committed to broad and unrestricted dissemination of knowledge. The spillovers principle developed as the

---

77. Nichols v. Universal Pictures Corp., 45 F.2d 119, 121 (2d Cir. 1930) (designating as "ideas" ineligible for copyright protection "Einstein's Doctrine of Relativity, or Darwin's theory of the Origin of Species").

78. *See* RONAN DEAZLEY, ON THE ORIGIN OF THE RIGHT TO COPY: CHARTING THE MOVEMENT OF COPYRIGHT LAW IN EIGHTEENTH CENTURY BRITAIN (1695–1775) 2, 221 (2004); LYMAN R. PATTERSON, COPYRIGHT IN HISTORICAL PERSPECTIVE 15, 43 (1968); MARK ROSE, AUTHORS AND OWNERS: THE INVENTION OF COPYRIGHT 12 (1993).

79. I EDWARD ARBER, A TRANSCRIPT OF THE REGISTER OF THE COMPANY OF STATIONERS, 1554–1640 A.D. at xxviii–xxxii (1876), available at *Primary Sources on Copyright (1450–1900)*, PRIMARY SOURCES ON COPYRIGHT (L. Bently & M. Kretschmer eds.), https://www.copyrighthistory.org/cam/tools/request/showRecord.php?id=record_uk_1557 [https://perma.cc/Y7SD-ZL4M].

80. Stationers' Company, London, *To the High Court of Parliament: The Humble Remonstrance of the Company of Stationers, London*, EARLY ENGLISH BOOKS ONLINE (1643), https://quod.lib.umich.edu/e/eebo2/A91370.0001.001/1:1?rgn=div1;view=fulltext [https://perma.cc/V3Y7-J33V].

81. 1710 8 Ann., c. 19, sec. 1 (Eng.).

82. *See* DEAZLEY, *supra* note 78, at 29.

83. 1710 8 Ann., c. 19, pmbl. (Eng.).

central mechanism for managing this tension by structurally limiting copyright's reach. Copyright, the principle assured, is limited to the making of copies and leaves free any knowledge or ideas.

The spillovers principle was most elaborately discussed and crisply developed in the public writings and official decisions surrounding the eighteenth century literary property debate — the struggle over the recognition of copyright as a common law property right.[84] Opponents of literary property often decried the dangers of knowledge "bound in such cobweb chains"[85] or of placing "manacles upon science."[86] This position was grounded in an understanding of the advancement of human knowledge and culture as a cumulative process. "The Learning of the present Age," one writer wrote, "may be considered as a vast Superstructure to the rearing of which the Geniusses of past Times have contributed their Proportion of Wit and Industry."[87] The response to this concern was a firm insistence that copyright does not apply to knowledge. Copyright, the argument went, is a narrow right to multiply copies that leaves "all the knowledge, which can be acquired from a contents of a book . . . free for every man's use," whether that knowledge is "mathematics, physic, husbandry," or even the skill of creating something new in the same genre as the protected work.[88] Much the same pattern — concerns over the circulation of knowledge responded to with firm assurance that copyright is a narrow right to make copies — was replicated in the American version of the literary property debate in the early nineteenth century.[89]

Today we commonly refer to the modern version of this legal structure as the idea/expression dichotomy.[90] But seeing the idea/expression

---

84. *See generally* MARK ROSE, *Battle of the Booksellers*, *in* AUTHORS AND OWNERS: THE INVENTION OF COPYRIGHT 67, 67–91 (1993); BRAD SHERMAN & LIONEL BENTLY, *Property In Mental Labour*, *in* THE MAKING OF MODERN INTELLECTUAL PROPERTY LAW: THE BRITISH EXPERIENCE, 1760–1911, 11, 11–42 (1999); DEAZLEY, *supra* note 78, at 115–28.

85. 17 COBBETT'S PARLIAMENTARY HISTORY OF ENGLAND 1001 (1813) (recording Lord Camden's opinion in Donaldson v. Becket, 1774).

86. Cary v. Kearsley [1802] 170 Eng. Rep. 680 (KB) (Lord Ellenborough, J.).

87. *An Enquiry into the Nature and Origin of Literary Property* 4–5 (1762), *reprinted in* HORACE WALPOLE'S POLITICAL TRACTS 1747–48 (Stephen Parks ed., 1974).

88. Millar v. Taylor [1769] 98 Eng. Rep. 216 (KB) (Willes, J.). For the last proposition, see *id.* ("[I]f, reading an epic poem, a man learns to make epic poems of his own; he is at liberty.").

89. *See* OREN BRACHA, OWNING IDEAS: THE INTELLECTUAL ORIGINS OF AMERICAN INTELLECTUAL PROPERTY, 1790–1909, at 143–45 (2016).

90. *See* WILLIAM F. PATRY, 2 PATRY ON COPYRIGHT § 4:31 (2018); STAFF OF H. COMM. ON THE JUDICIARY, 87TH CONG., COPYRIGHT LAW REVISION: REPORT OF THE REGISTER OF COPYRIGHTS ON THE GENERAL REVISION OF THE U.S. COPYRIGHT LAW 3 (1961) ("Copyright does not preclude others from using the ideas or information revealed by the author's work. It pertains to the literary, musical, graphic, or artistic form in which the author expresses intellectual concepts. It enables him to prevent others from reproducing his individual expression without his consent. But anyone is free to create his own expression of the same concepts, or to make practical use of them, as long as he does not copy the author's form of expression.").

dichotomy as a technical legal rule undersells its significance. The dichotomy embodies the spillovers principle as a constitutive, foundational principle of the field, grounded in multiple doctrinal structures. The broad principle is latent. There is no Section 1 of the Copyright Act stating it. Nevertheless, two features make the spillovers principle a fundamental one. First, it cuts across many specific rules, giving them a common coherent meaning and a unifying purpose grounded in a central concern of the field. Second, it is a general structural feature, rather than a negotiated policy call that instructs decisionmakers to optimize the application of the rules on a case-by-case basis. As Matthew Sag aptly puts it, what is at issue here is not "just some ad hoc compromise, or a shifting equilibrium," but rather "a deep fundamental structure that revolves around the protection of original expression."[91]

To limit copyright's toll on the development of knowledge and culture, the spillovers principle structurally restricts copyright's exclusionary effect on two levels. On the primary level, copyright is tightly restricted to the domain of expressive forms. Any other informational subject matter, even if it comprises the primary source of value of the relevant information good, is outside copyright's purview.[92] And the principle applies both to information goods that are completely non-expressive and to non-expressive elements bundled with expressive ones in a single good. The domain aspect of the spillovers principle is implemented in a litany of doctrines,[93] including the idea/expression dichotomy that prevents protection of both knowledge (conceptual or factual)[94] and high abstraction level expressive elements;[95] the exclusion of functional subject matter;[96] the "scenes a faire" doctrine that prevents protection of expressive elements that hold a dominant status

91. Sag, *supra* note 46, at 303.

92. *See* Molly Shaffer Van Houweling, *The Freedom to Extract in Copyright Law*, 103 N.C. L. REV. (forthcoming 2025) (manuscript at 6) (on file with authors) ("It is a foundational principle of copyright law that protection attaches only to the expression embodied in copyrighted works, not to the underlying substance conveyed by that expression."). Certain non-expressive information goods or elements can be protected by other legal regimes under their relevant requirements and terms. Information embodying useful inventions, for example, may be protected by a patent. Many other valuable aspects of informational works simply fall into the public domain.

93. *See id.* at 13 (describing how a "number of doctrines in copyright law" recognize that "extractive use of some expression can be necessary to fully vindicate" the freedom to extract and use non-expressive elements).

94. 17 U.S.C. § 102(b) (defining ineligible subject matter including "idea . . . concept, principle, or discovery").

95. *Id.*; Nichols v. Universal Pictures Corp., 45 F.2d 119, 122 (2d Cir. 1930) (articulating the abstraction test and observing that "too generalized an abstraction" is unprotectable as being "ideas").

96. See Baker v. Selden, 101 U.S. 99, 104 (1879); 17 U.S.C. § 102(b) (defining ineligible subject matter including "procedure, process, system, method of operation").

within a genre of expression;[97] the supporting doctrine of merger;[98] and even the creativity prong of the originality requirement.[99]

On the secondary level, even expressive subject matter receives only protection that is limited in scope. The scope dimension of the spillovers principle means that rather than a plenary power to exclude, copyright confers a well-defined and limited exclusionary power on owners. This dimension of the principle too is implemented in a series of doctrines: the infringement test that restricts actionable infringement to a zone of substantially similar copies;[100] enumerated entitlements that circumscribe the right to exclude to a closed list of specific activities rather than any valuable use of works;[101] and at the back end, various exemptions and carve-outs, the most important of which is the fair use defense.[102]

The spillovers principle and the elaborate doctrinal structure that implements it are grounded in the purpose of modern copyright and the basic dynamics of producing and using information goods that underlie it. The fundamental tenet is that copyright is not about full internalization of value by producers of information goods.[103] Whatever the merits of full internalization, or a so-called absolute right to exclude, with respect to property rights in other resources, this purpose is simply not applicable to copyright.[104]

The economics of expressive goods involve a dynamic side, relating to their production, and a static one, relating to their use or consumption. Dynamically, supporting creation requires a level of

---

97. Schwarz v. Universal Pictures Co., 85 F. Supp. 270, 275 (S.D. Cal. 1945).

98. Under the merger doctrine, when functional subject matter is merged with expression, it is unprotectable by copyright. *See Baker*, 101 U.S. at 104–105 (ruling that a useful "art" is unprotectable by copyright even when its exercise "correspond[s] more closely" with using specific expressive materials).

99. Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 345–46 (1991) (describing originality as requiring independent creation and a modicum of creativity). The creativity prong is best understood as an additional partially redundant filter for expressive subject matter. Oren Bracha & John M. Golden, *Redundancy and Anti-Redundancy in Copyright*, 51 CONN. L. REV. 247, 290 (2019).

100. *See* Arnstein v. Porter, 154 F.2d 464, 473 (2d Cir. 1946).

101. 17 U.S.C. § 106.

102. *See* 17 U.S.C. §§ 107–22.

103. *See* Jessica D. Litman, *Fetishizing Copies*, *in* COPYRIGHT IN AN AGE OF LIMITATIONS AND EXCEPTIONS 79 (Ruth Okediji ed., 2017) (observing that "[c]opyright owners are not entitled to control many valuable uses of their works").

104. The notion of absolute property rights is conceptually incoherent in general. *See* Talha Syed & Anna di Robilant, *Property's Building Blocks: Hohfeld in Europe and Beyond*, *in* THE LEGACY OF WESLEY HOHFELD: EDITED MAJOR WORKS, SELECT PERSONAL PAPERS, AND ORIGINAL COMMENTARIES 229 (Shyamkrishna Balganesh, Ted Sichelman & Henry E. Smith eds., 2022). Full internalization is not a conceptually incoherent notion, but an extremely unattractive goal especially with respect to information goods. *See* Julie E. Cohen, *Lochner in Cyberspace: The New Economic Orthodoxy of "Rights Management*,*"* 97 MICH. L. REV. 462, 502 (1998); Mark A. Lemley, *Property, Intellectual Property, and Free Riding*, 83 TEX. L. REV. 1031, 1037–38 (2005); Oren Bracha, *Give Us Back Our Tragedy: Nonrivalry in Intellectual Property Law and Policy*, 19 THEORETICAL INQUIRIES L. 633, 648 (2018).

exclusion sufficient to enable the producer to price at a level that covers production cost.[105] Any additional iota of copyright exclusionary power achieves nothing by way of enabling creation and comes with the dual negative effect of restricting access to works and erecting barriers to downstream development of other works (aka "deadweight loss").[106]

What often escapes notice is that unlike other contexts, proprietary power to exclude serves no purpose with respect to the static aspect of coordinating the use of existing expressive works.[107] Such works, like most information goods, are nonrival, which means that the use or enjoyment by one person does not decrease the ability of others to do so.[108] There is no tragedy of the commons with respect to information goods.[109] The upshot is that no governance mechanism of exclusion and coordinating use — proprietary or otherwise — is necessary, and that any mechanism of that sort will be a pure negative on the static side.[110] "Ex post" theories of copyright, that are supposedly based on grounds related to static use rather than dynamic production incentives, fail to change this conclusion.[111] On close examination, such theories are either responses to concerns of dynamic production or unpersuasive attempts to refute the assumption of nonrivalry of expressive goods.[112]

The irrelevance of full internalization is not an unfortunate side effect or a second-best outcome due to transaction costs that frustrate coordination of uses via market transfers.[113] With respect to strongly nonrival expressive goods, full internalization is simply not a goal even in a fantastical frictionless world of zero transaction costs.[114] With grounding in neither dynamic support of production nor static coordination of use, full internalization serves no purpose.

This irrelevance of full internalization holds, given any plausible consequence-oriented incentive basis of copyright: utilitarianism, economic efficiency, or a variant of democratic theories.[115] Neither does

---

105. Stan J. Liebowitz, *Is Efficient Copyright a Reasonable Goal?*, 79 GEO. WASH. L. REV. 1692, 1698 (2011) (explaining that efficient copyright should "last for just long enough that the profits being earned in the publishing market would be exactly sufficient to cover the cost of creation").

106. *See* Bracha, *supra* note 104, at 662.

107. *See id.* at 641.

108. *See* Bracha, *supra* note 104, at 634; RICHARD CORNES & TODD SANDLER, THE THEORY OF EXTERNALITIES, PUBLIC GOODS, AND CLUB GOODS 6 (1996).

109. *See* Bracha, *supra* note 104, at 634.

110. *Id.* at 641.

111. Mark A. Lemley, *Ex Ante and Ex Post Justifications for Intellectual Property*, 71 U. CHI. L. REV. 129, 129 (2004).

112. Bracha, *supra* note 104, at 658.

113. *Id.* at 647.

114. *Id.* at 647–48.

115. *See* Oren Bracha & Talha Syed, *Beyond Efficiency: Consequence-Sensitive Theories of Copyright*, 29 BERKELEY TECH. L.J. 229, 244–47 (2014) (discussing the structural similarity between economic efficiency theory of copyright and competing consequence-sensitive theories).

the purpose of just deserts to creators, whether on its own in the guise of a natural rights justification or as a component of another theory, require full internalization. While it is plausible to argue that creators morally deserve reward for their effort and sacrifice, even beyond covering their cost,[116] it is much less plausible to assume that they can take credit and therefore have a moral claim for the full social value of their creation.[117]

In short, spillovers — a right to exclude structurally restricted in both domain and scope that results in much of the social value of expressive works being externalized — is a feature, not a bug. Copyright is about spillovers, not full internalization.

With a firm understanding of the spillovers principle in mind — specifically its domain aspect that strictly limits copyright to expression — it becomes clear that the unorthodox arguments of GenAI infringement are doomed to fail. Such arguments, whether upstream or downstream, fly in the face of the spillovers principle and the basic structure of modern copyright.

IV. UPSTREAM: TRAINING COPIES

In the wake of the rise of GenAI, the argument that reproduction of a copyrighted work strictly as part of the training process is in itself infringement has been gathering momentum.[118] As Mark Lemley and Bryan Casey explain, the implications of this argument go far beyond the context of GenAI expressive production.[119] Any system based on machine learning, whatever its output or purpose, requires training with large amounts of materials. For a system to analyze it, the material must be reproduced in a digital form. Given the ubiquity of copyright, the extremely low threshold for its validity, and the fact that rights attach

116. Liebowitz, *supra* note 105, at 1692 (discussing "fairness" concerns with limiting creators' rights to those strictly necessary to induce production).

117. Bracha & Syed, *supra* note 115, at 295–96 (discussing the distributive concern of fair compensation to creators); *see* Shubha Ghosh, *The Merits of Ownership*, 15 HARV. J.L. & TECH. 453, 477 (2002) (asking "[w]hat is a reasonable rate of return" to creators under intellectual property rights).

118. *See, e.g.*, Complaint at 3, 8, 12–19, Getty Images v. Stability AI, 23-cv-00135 (D. Del. Feb. 3, 2023); CHRISTOPHER T. ZIRPOLI, CONG. RSCH. SERV., LSB10922, GENERATIVE ARTIFICIAL INTELLIGENCE AND COPYRIGHT LAW 3, 5 (2023) (observing that the "training process involves making digital copies of existing works" which carries a "risk of copyright infringement"); Enrico Bonadio, Plamen Dinev & Luke McDonagh, *Can Artificial Intelligence Infringe Copyright? Some Reflections*, *in* RESEARCH HANDBOOK ON INTELLECTUAL PROPERTY AND ARTIFICIAL INTELLIGENCE 247 (Ryan Abbott ed., 2022) [hereinafter "Handbook on IP & AI"] (observing that training copies "may violate the right to reproduction").

119. *See* Mark A. Lemley & Bryan Casey, *Fair Learning*, 99 TEX. L. REV. 745–46 (2021); *see also* Benjamin Sobel, *A Taxonomy of Training Data: Disentangling the Mismatched Rights, Remedies, and Rationales for Restricting Machine Learning*, *in* ARTIFICIAL INTELLIGENCE AND INTELLECTUAL PROPERTY 1, 6–7 (Reto Hilty, Jyh-An Lee & Kung-Chung Liu eds., 2021).

automatically at creation and are hard to opt out of, if reproduction in training were recognized as violating copyright, infringement would be omnipresent.[120] Use in training of expressive GenAI is merely a subset of the enormous terrain threatened by the shadow of this infringement argument. The vast coverage makes the question urgent: is reproduction in training copies infringing?

A. Non-Expressive Extraction and Learning

The infringement argument is deceptively simple. Its heart is the assertion that a copy is a copy. More specifically, use of a work in a training set requires making a digital copy of the work.[121] The digital copy is itself reproduction and therefore constitutes copyright infringement, as long as the material is under copyright (which it almost always is).[122] No matter that no human eye or ear will ever experience the work from its new physical embodiment. No matter that the only purpose of the reproduction is the extraction of metadata necessary for the machine learning process, i.e. for allowing the machine to acquire the capacity to generate new and non-infringing works. No matter what kind of new materials the machine produces, and, indeed, no matter if it is thrown into a ditch before producing any materials. The training digital copy is infringement, period. There is a remarkably broad agreement on this proposition.[123] The main debate centers on whether, given the prima-facie infringement, special circumstances are present that justify exempting this kind of reproduction as fair use.[124]

---

120. Lemley & Casey, *supra* note 119, at 754–55.

121. *But see* Sobel, *supra* note 119, at 228 (suggesting that "technological progress may obviate the need to fix training data at all").

122. *See* 17 U.S.C. § 106(1) (giving copyright owners the right "to reproduce the copyrighted work in copies"); *see also* 17 U.S.C. § 101 (defining "copies" as "material objects . . . in which a work is fixed").

123. There are at least two notable exceptions to this consensus: BJ Ard, *Copyright's Latent Space: Generative AI and the Limits of Fair Use*, 110 CORNELL L. REV. __, 36 (forthcoming 2025) (manuscript at 68) (on file with authors) (arguing that "fair use, particularly the emphasis on transformative purpose, struggle to map onto the realities of how these AI models operate"); Amanda Levendowski, *How Copyright Law Can Fix Artificial Intelligence's Implicit Bias Problem*, 93 WASH. L. REV. 579, 595–96 (2018) (while focusing mainly on the fair use analysis, observing that "[c]ourts have also yet to confront whether unauthorized copies made for training AI are necessarily infringing copies" and suggesting other possible reasons for non-infringement).

124. *See, e.g.*, Andrew W. Torrance & Bill Tomlinson, *Training is Everything: Artificial Intelligence, Copyright, and "Fair Training,"* 128 DICK. L. REV. 233, 233 (2023) (surveying arguments for and against fair use); Lemley & Casey, *supra* note 119, at 759; Bonadio et al., *supra* note 118, at 247–52; Samuelson, *supra* note 51, at 159–61 (surveying the possible fair use analysis of claims against training copies in ongoing lawsuits); Jessica L. Gillotte, *Copyright Infringement in AI-Generated Artworks*, 53 U.C. DAVIS L. REV. 2655, 2680–84 (2020); Daryl Lim, *AP & IP Innovation: Creativity in An Age of Accelerated Change*, 52 AKRON L. REV. 813, 847 (2018); James Grimmelmann, *Copyright for Literate Robots,* 101 IOWA L. REV. 657, 661–65 (2016).

This copy-fundamentalism, however, violates the spillovers principle, and is infected with confused physicalism. The spillovers principle, recall, mandates that copyright is strictly limited to the domain of expressive forms.[125] A copyright owner receives a right to exclude others from engaging in certain activities pertaining to enjoying the use value of her expression qua expression. All other aspects of the information good in which the expression is embedded, no matter how valuable, are outside the domain of copyright. These non-expressive aspects include any knowledge communicated by the information good, but also the meta-knowledge required for learning how to produce expression.

Consider the following example. A historical literary novel may directly communicate certain knowledge via its content — for example, a factual recap of the events of the Napoleonic wars. The novel may also be the source of a meta-knowledge that is not its communicative content: by studying the expression in the novel (and others like it), one may obtain the knowledge, that is, learn the techniques and acquire the skills, of generating an expressive good of a similar kind but with different expressive forms. Both kinds of knowledge — the communicative content of a work and the meta-knowledge of expressive skills — have always been outside copyright's domain.[126] Learning how to produce other works by extracting meta-information from existing ones is at the heart of the spillovers concern of placing knowledge and its cumulative accretion beyond the reach of copyright.[127] So much so that in traditional copyright contexts, arguments for excluding others from learning are almost non-existent; arguing that my copyright allows me to exclude you from the valuable use of my novel by way of learning how to write novels would be seen as outlandish. This holds even if you use that meta-knowledge to write your own historical novel that competes with mine in the market for expressive goods. As long as the competing product did not appropriate the expressive forms from the original, market competition is irrelevant. Again, to argue that you infringe because you learned how to write novels by reading mine, which now enables you to produce a different novel that competes with mine, is so outlandish that no one makes the argument.

The analysis of infringement by way of machine learning is identical except, of course, that machine, unlike human, learning requires background reproduction, at least for now.[128] The key point, however, is that the learning-copy changes nothing with respect to the applicable purposes and concerns. Making a digital copy is simply an essential step in how machines work. There is no machine learning with no

125. *See supra* text accompanying notes 92–99.
126. *See supra* text accompanying notes 87–89.
127. *See supra* text accompanying notes 88–89.
128. *See supra* text accompanying notes 55–62.

learning-copy. The learning-copy is completely incidental to, and is used strictly for, the machine learning process. No one ever enjoys the work's expressive value through the learning-copy. In short, machine learning is a new technological equivalent of the process of extracting meta-knowledge out of an expressive good, as in the case of learning how to write a novel by reading existing literature. This kind of knowledge has always been placed beyond copyright's domain, and the existence of a training copy, which is completely incidental to the learning process, changes nothing in this analysis.

Nor does it matter if the expressive product generated by the machine causes "market harm" by competing with the copyrighted work.[129] All expressive creation learns from existing works and then results in "market harm" by competing with those works. But this sort of competitive market effect empowered by learning has always been regarded as a boon rather than a fault.[130] Just as in the traditional case of an independently-created novel, a market effect that is not traceable to reusing protectable expression in a competing expressive product is irrelevant. It may be a market effect, but it is not a market "harm" cognizable by copyright. The relevant element on which copyright liability depends is not mere market-encroachment, but market-encroachment caused by enabling access to the use value of protected expression.

B. The Physicalist Fallacy

But what about "a copy is a copy?" Can we really ignore the fact that in the case of GenAI training, unlike more traditional learning, a new physical embodiment of the work is produced? Doesn't the physical copy make all the difference? The answer is that insisting on extending copyright's exclusion power to learning on the sole basis of the presence of a physical embodiment is a deep misunderstanding of the domain of copyright as a field of *intellectual* property.[131] To state the obvious: copyright's object of property is not a physical phenomenon

129. Some argue that reproduction in machine training may count as copyright infringement as long as the final product is "market encroaching" even if it does not incorporate any of the expressive forms of the copyrighted work. *See* Sobel, *supra* note 119, at 231–33; *see also* Benjamin L. W. Sobel, *Artificial Intelligence's Fair Use Crisis*, 41 COLUM. J.L. & ARTS 45, 77–79 (2017) (arguing that market-encroaching uses, even if the competing product does not copy expression, may not be entitled to the fair use privilege).

130. *See* Sega v. Accolade, 977 F.2d 1510, 1523 (9th Cir. 1992) (explaining with respect to the study of unprotected elements used to develop independent competing works that "[i]t is precisely this growth in creative expression, based on the dissemination of other creative works and the unprotected ideas contained in those works, that the Copyright Act was intended to promote").

131. *See* Talha Syed, *Reconstructing Patent Eligibility*, 70 AM. U. L. REV. 1937, 1949–53 (2021) (developing the point that the objects of both copyright and patent protection need to be conceived in thoroughly dephysicalized ways as, respectively, intangible forms of expression and intangible spaces of applied knowledge).

of any kind, but rather a purely informational good. Specifically, copyright's proper subject matter is a particular kind of information, i.e. expressive forms.[132] The full implications of this subject-matter domain of the field are sometimes less obvious. Copyright domain's focus on expression means that the basic purpose of the field is grounded in the production and use dynamics of expression and expression alone. Physical facts — whether the making of physical objects, their display, or transfer of their possession — are never relevant in themselves. These physical facts are relevant only to the extent they involve in some way access to the use value of protected expression.

Making a new physical copy when the expression embodied in it will be experienced by no one is no more relevant for copyright than using an existing copy as a doorstop.[133] A moon-lander that produces on the dark side of a moon a printout of a poem not to be seen by anyone — not even by a video transmission — is a farfetched example, but it demonstrates the physicalist fallacy of basing copyright analysis on physical activities detached from access to expressive use value.[134] Insisting that the mere physical fact of reproduction is a sufficient condition for triggering the relevance of copyright is a form of fetishism.[135] It is a prime example of what Jessica Litman aptly called "copy-fetish."[136] This position maintains that physical objects in themselves have significance, perhaps even mysterious powers to invoke the relevance of copyright. Instead, what really matters is how and whether these physical objects relate to relevant human interests and activities, in the case of copyright, interests and activities related to the production and use of the value of expression.

There is no shortage of specific policy reasons to avoid regarding training copies as infringing. Unhindered access to a broader training set improves the results of AI systems.[137] Access to broad and diverse training data is also necessary, even if not sufficient, for ameliorating concerns about biases and fairness in AI output, as well as concerns

---

132. *See supra* text accompanying notes 92–99.

133. *See* ABRAHAM DRASSINOWER, WHAT'S WRONG WITH COPYRIGHT? 87 (2015) (explaining that "merely technical reproduction incidental to the operation of digital technology cannot give rise to liability"); Abraham Drassinower, *Remarks on Technological Neutrality in Copyright Law as a Subject Matter Problem: Lessons from Canada*, 81 CAMBRIDGE L.J. 50, 56 (2022).

134. If one seeks a less farfetched example of the work of the physicalist fallacy in copyright, the prime exhibit is the case law under which computer Random Access Memory copies constitute infringing reproduction. *See, e.g.*, MAI Sys. Corp. v. Peak Comput., Inc., 991 F.2d 511, 511 (9th Cir. 1994). *But see* Litman, *supra* note 103, at 80–86 (critiquing this case law along the lines of physicalism).

135. *See* KARL MARX, CAPITAL 163–77 (Ben Fowkes trans., Penguin Books 2004) (1867) (discussing "The Fetishism of the Commodity").

136. Litman, *supra* note 103, at 76 (defining "copy-fetish" as the idea that any physical reproduction is infringing, "whether or not anyone will ever see the copy").

137. *See* Lemley & Casey, *supra* note 119, at 770.

about concentrated control of such systems.[138] Finally, given the vast amounts of material used in training sets and the multitude of rights and owners involved, licensing of rights is sure to be greatly incumbered, sometimes fatally so, by transaction costs. The predictable outcome is crippling the development of AI systems accompanied by little compensation for copyright owners (for transactions that would not take place).[139]

While these specific policy reasons are important, there are two deeper, related reasons for adhering to the spillovers principle's exclusion of non-expressive uses from the domain of copyright. First, the spillovers principle is a foundational structural feature of copyright. It is grounded in the field's purpose and deepest commitments, but it is not a shifting policy equilibrium that depends on case-specific optimization.[140] Training copies are beyond copyright's reach before ever invoking context-specific policies. Mere physical reproduction, delinked from enjoyment of the expressive value of a work and completely incidental to accessing unprotected meta-information, is categorically beyond copyright's domain.

Second, the structural character of the principle is reflected in its doctrinal instantiation. Attempts to protect non-expressive aspects of information goods should be blocked at the very threshold of subject matter rules, not at the back-end of a fair use exemption.[141] An activity that is purely about learning meta-knowledge rather than appropriating the use value of expression involves non-copyrightable subject matter, or in copyright law's terms, mere "ideas."[142]

Even more technically, if one wanted to (wrongly) insist on the physical fact that the expression is reproduced in the training copy, the merger doctrine would dismiss this argument. The merger doctrine, that operates as a crucial adjunct to subject matter rules, provides that in cases where using expression is indispensable for accessing and using non-protectable elements of a work, the expression and the unprotectable element merge.[143] In such cases, the use is allowed, but only to the extent necessary for accessing the unprotectable material.[144] In the case of completely incidental training copies, accessing the unprotectable meta-knowledge necessitates (at least in a narrow physicalist sense)

138. *See* Levendowski, *supra* note 123, at 592; Lemley & Casey, *supra* note 119, at 771–72.

139. *See* Lemley & Casey, *supra* note 119, at 770–71.

140. *See* Sag, *supra* note 46, at 303.

141. *See* Drassinower, *supra* note 133, at 59.

142. *See* 17 U.S.C. § 102(a)–(b).

143. *See* Baker v. Selden, 101 U.S. 99, 104–05 (1879).

144. *Id.*

reproducing the expression.[145] The latter therefore merges with the former and its copying is outside the domain of copyright.[146] In plain words, the reproduction of the physical patterns representing the work in the belly of the machine is a mere physical incident that inevitably attaches to a permissible learning process when done in digital rather than analog.

C. Subject Matter, not Fair Use

Unfortunately, the case law, including a firm line of precedents that are friendly to allowing the making of training copies — has taken a wrong turn on the doctrinal front. This wrong turn placed the burden of exempting non-expressive copies on the too slender shoulders of the fair use doctrine. Courts premised central decisions, especially those involving incidental reproductions of copyrighted works by digital technology, on applications of fair use. Even if they ultimately exempted the reproduction, these courts assumed that copyrightable subject matter was implicated, and that prima facie infringement had occurred, and only then proceeded to ask whether the use was fair.

The seminal decision of this type is *Sega v. Accolade*.[147] The case involved a defendant who wished to develop independently-created video games for Sega's game console without Sega's permission. To achieve this, the defendant had to obtain access to the communication protocols of the console that constituted functional information unprotectable by copyright.[148] However, the only way of obtaining this information was creating intermediary copies of the copyrighted code of Sega's games and then extracting the functional specifications by reverse engineering. The crucial feature of the case was that the reproduction of the computer code was entirely incidental to extracting the unprotectable information.[149] No one enjoyed the *expressive* value of the reproduced video games. The Ninth Circuit's critical first step in analyzing the case was concluding that the non-expressive copying of the code constituted prima-facie infringement.[150] Its reasoning was that

---

145. In fact, conceding that training copies involve reproducing expression *in the sense relevant for copyright* and that therefore the merger doctrine is necessary to avoid liability is already a stretch. One may call the physical object an embodiment of the expression in some technological or ontological sense, but it is not in the sense relevant for copyright's subject matter, that is, a physical object from which someone will enjoy the expressive value of the work. *See supra* text accompanying notes 133–136.

146. This does not hold, of course, with respect to either further uses of the training copies which are not merely incidental to machine learning and involve access to the expressive value of the work, or further reproductions of the work or substantially similar versions in the system's output. *See* Sobel, *supra* note 119, at 65.

147. 977 F.2d 1510 (9th Cir. 1992).

148. *Id.* at 1514–15.

149. *Id.* at 1532.

150. *Id.* at 1518.

a copy is a copy.[151] The court based its decision on "the plain language of the Act": the fact that a copyright owner is granted a clear right to reproduce the work in copies and that the defendant's actions fell squarely within the statutory definition of fixing a work in a material copy.[152] The court then found that defendant's reproduction was, nevertheless, exempted by the fair use defense.[153]

Sega set the pattern that was followed by virtually all subsequent cases. A line of cases involving intermediary or non-expressive copies where no one accessed the expressive value of the work concluded that such copies were non-infringing.[154] However, as in *Sega*, the courts in these cases assumed prima-facie infringement and proceeded to apply the fair use defense.[155] In virtually all of these cases, the most important of which pertained to the full digital reproduction of numerous books as part of the Google Books search engine, courts found that the reproduction was exempted as fair use.[156] Since *Sega*, this has been the uniform pattern in both court decisions and scholarship: a consensual acceptance that making non-expressive copies constitutes actionable reproduction, followed by a clear trend to exempt under fair use.[157]

The result in *Sega* is correct, but its reasoning that has dominated the legal terrain ever since is flawed. Non-expressive copies involve no enjoyment of any expression qua expression. Reproduction in the technical sense is a mere physical fact which has nothing to do with copyright's domain and purpose.[158] The only information good whose use value is enjoyed is either completely different expressive goods, as in the new video games in *Sega*, or unprotectable subject matter such as functional elements, meta-information, or knowledge. As a result, the copying does not involve any copyrightable subject matter and should be found non-infringing long before ever reaching the fair use question.[159]

---

151. *Id.*

152. *Id.* at 1518–19.

153. *Id.* at 1522–28.

154. Mathew Sag has appropriately dubbed such cases "nonexpressive" uses. Matthew Sag, *Copyright and Copy-reliant Technology*, 103 NW. UNIV. L. REV. 1607, 1624 (2009).

155. *See, e.g.*, Sony Comput. Ent., Inc. v. Connectix Corp., 203 F.3d 596, 602 (9th Cir. 2000); A.V. ex rel. Vanderhye v. iParadigms, LLC, 562 F.3d 630, 645 (4th Cir. 2009); Authors Guild, Inc. v. HathiTrust, 755 F.3d 87, 101 (2d Cir. 2014).

156. Authors Guild v. Google, Inc., 804 F.3d 202, 225 (2d Cir. 2015) (holding that making digital copies of copyrighted books used only to facilitate digital searches of the books' texts was presumptively infringing but exempted as fair use).

157. *But see* Drassinower, *supra* note 133, at 56.

158. *See supra* text accompanying notes 133–134.

159. Lemley and Casey see clearly that copying non-expressive subject matter, even when a verbatim non-expressive copy is made, does not involve copyrightable subject matter under traditional subject matter doctrines. Nonetheless, they conclude that the non-expressive incidental copy, one whose making is inescapable with digital technology, inevitably violates the reproduction entitlement. They then fall into the general pattern of arguing that such

Nor does the *Sega* court's "plain meaning" statutory interpretation reasoning change this conclusion.[160] To be sure, the statutory definitions of the reproduction right and of fixation in copies do not exclude non-expressive copying.[161] But to conclude that this makes all reproduction prima facie infringement, subject only to fair use, is a non sequitur. The statutory definitions chart the boundaries of the right to exclude by specifying the relevant activity encompassed by it. In the modern statute, the relevant definition of reproduction includes making an object from which the work is perceivable only indirectly, such as microfilm or digital files.[162] But the definition says nothing about additional limitations on liability imposed by other copyright principles and rules. It does not follow from a definition of reproduction that encompasses making a copy from which a work is only indirectly perceivable that all reproduction is actionable, even if the expression embodied in the copy will never be enjoyed by humans at all, directly or indirectly.[163]

Following a long tradition, fundamental limiting principles of copyright are often grounded in the case law and have little reflection in the statutory text. No one questions that actions that fall within the statutory definition of reproduction but do not satisfy the infringement test of substantial similarity of expression[164] or general principles of "volition"[165] are not infringing. This is so even though neither the infringement test nor "volition" has any grounding in statutory language. Exemptions aside, that an activity is within the statutory definitions of the exclusive rights is simply an insufficient condition for infringement. Just as a reproduction that fails to satisfy the infringement test is non-

---

reproduction should be treated as fair use. In other words, despite seeing clearly the subject matter issue, the authors, perhaps simply due to bowing to existing case law, fall in the trap of physicalism. *See* Lemley & Casey, *supra* note 119, at 772–73, 775.

160. Sega v. Accolade, 977 F.2d 1510, 1518 (9th Cir. 1992).

161. *See* 17 U.S.C. § 101 (defining "copies" as "material objects . . . in which a work is fixed by any method now known or later developed, and from which the work can be perceived, reproduced, or otherwise communicated, either directly or with the aid of a machine or device" and providing that a work is "fixed" "when its embodiment in a copy or phonorecord . . . is sufficiently permanent or stable to permit it to be perceived, reproduced, or otherwise communicated for a period of more than transitory duration").

162. *Id.*

163. For an argument that the historical legislative expansion of the legal definition of a copy necessitates regarding all reproduction as prima facie infringement, see Sag, *supra* note 46, at 308–09. To see why this is a non sequitur, consider the case of reproducing a musical composition in the form of a perforated role of a player piano that is designed to play the tune. Historically, this was considered not to be reproduction because the work was not directly perceivable. *See* White-Smith v. Apollo, 209 U.S. 1, 18 (1908). Today, this is no longer the case. But even today, there is a fundamental distinction between this case, where the whole purpose of the reproduction is to make the expressive value of the work ultimately available to humans, albeit with the aid of a device, and cases where no human will enjoy such expressive value via the physical copy at all.

164. *See, e.g.*, Arnstein v. Porter, 154 F.2d 464, 468 (1946).

165. Cartoon Network LP, LLLP v. CSC Holdings, Inc., 536 F.3d 121, 131 (2d Cir. 2008).

infringing, so too is a reproduction that fails to be relevant for any copyrightable subject matter.[166]

None of the analysis above means that copyright infringement requires a plaintiff to specifically prove as part of the prima facie case that someone actually accessed the expressive use value of the work associated with a particular infringing activity. There are many examples of infringement by actions that involve potential human enjoyment of the protected expression even if such enjoyment has not yet occurred or been established. A "bootlegger" caught with a massive stock of copies he illegally reproduced does not get to escape because no one enjoyed the expressive value of these copies yet. Indeed, copyright's most fundamental entitlement — the right of reproduction — presupposes that no establishment of actual access or consumption is necessary.[167] Someone who made an unauthorized copy is an infringer whether the plaintiff can show that anyone consumed the work via the copy or not. The same is true of the distribution right, that goes one step further toward requiring making an illicit copy available to others, but stops short of demanding actual consumption of the work.[168] The prophylactic logic of these entitlements is clear: if plaintiffs were required to establish actual enjoyment of expressive value, the evidentiary difficulties and cost would be enormous, copyright would be remarkably ineffective in practice and its purpose would be frustrated.

But this prophylactic logic does not apply to cases where structurally and in principle the relevant action does not involve human enjoyment of the expressive value of the work at all. Copies made in the belly of the machine for purposes of GenAI training or any other extraction of non-protectable information are not meant for expressive enjoyment by any human ear or eye, not even down the road.[169] It is not merely a matter of whether access to their expressive value has occurred yet or whether access can be proved. The point is that such access to the expressive value is not part of the purpose or the structural features of the process or action involved. As a result, the acts of reproduction are mere physical facts, rather than a case that falls within copyright's expressive domain even if the final stage of expressive consumption has not yet occurred.

One may be inclined to think that the different grounds for reaching a result of no infringement are immaterial: a dry and pointless lawyerly insistence on formal distinctions devoid of substance. Why should we care if certain acts are non-infringing because no copyrightable subject matter is implicated or on fair use grounds? This would be wrong.

166. Moreover, unlike the infringement test, subject matter principles do have grounding in the statutory text. 17 U.S.C. § 102(b).

167. 17 U.S.C. § 106(1).

168. 17 U.S.C. § 106(2).

169. *See supra* text accompanying notes 57–62.

There are significant implications to the alternative legal grounds, both practical and conceptual.

Practically, some central features of fair use make it inferior in performing the task of allowing non-expressive uses. Courts treat fair use as an affirmative defense and lay the burden for its establishment on defendants.[170] Furthermore, fair use, with its four-factor structure, is an extremely open-ended standard whose application is fact and legal analysis intensive.[171] This is true despite some mitigating effects of precedent and pattern formation over time.[172] This, in turn, makes low-cost procedural routes for early resolution of disputes, such as motion to dismiss and summary judgment, less available.[173] The upshot is two-fold. First, because fair use is a notoriously unpredictable and manipulation-prone legal rule,[174] small contextual differences may result in different outcomes or at least open the door to distinctions and challenges.[175] Unpredictability breeds risk and with it chilling effects on the activities of potential users.[176] Second, because fair use is a privilege that is very expensive to take advantage of — establishing fair use may often involve protracted and expensive legal proceedings.[177] Consequently, those actors who have superior resources and sophistication will enjoy the privilege disproportionally.

Even more importantly, fair use is an ill-fit for conceptual reasons — it fails to capture the substantive reason why training copies do not infringe. Under the spillovers principle, such copies do not infringe because the sine qua non of copyrightable subject matter — access to the expressive value of the work — is not implicated.[178] That is the business of subject matter rules that regulate the front entrance to copyright's domain. Fair use, by contrast, is a back-end doctrine, twice removed. Its proper province is regulating copyright's scope, and even

---

170. *E.g.*, Campbell v. Acuff–Rose Music, Inc., 510 U.S. 569, 590 (1994) (observing that "fair use is an affirmative defense"). This is notwithstanding loud scholarly protests. *See, e.g.*, Lydia Pallas Loren, *Fair Use: An Affirmative Defense?*, 90 WASH. L. REV. 685, 688 (2015) (arguing that the Supreme Court "should conclude that fair use is not an affirmative defense but is a mere defense").

171. Bracha & Golden, *supra* note 99, at 267–68.

172. *See* Barton Beebe, *An Empirical Study of U.S. Copyright Fair Use Opinions*, 1978–2005, 156 U. PA. L. REV. 549, 556 (2008); Pamela Samuelson, *Unbundling Fair Uses*, 77 FORDHAM L. REV. 2537, 2603 (2009); Neil Weinstock Netanel, *Making Sense of Fair Use*, 15 LEWIS & CLARK L. REV. 715, 729 (2011).

173. Bracha & Golden, *supra* note 99, at 267–69; Oren Bracha, *Not De Minimis: (Improper) Appropriation in Copyright*, 68 AM. U. L. REV. 139, 197–200 (2018) [hereinafter Bracha, *Not De Minimis*].

174. Pierre N. Leval, *Toward a Fair Use Standard*, 103 HARV. L. REV. 1105, 1132 (1990) (Whether copying "will pass the fair use test is difficult to predict. It depends on widely varying perceptions held by different judges.").

175. Lemley & Casey, *supra* note 119, at 763; *see also* Ard, *supra* note 123, at 36.

176. James Gibson, *Risk Aversion and Rights Accretion in Intellectual Property Law*, 116 YALE L.J. 882, 882 (2007).

177. *See* Bracha, *Not De Minimis*, *supra* note 173, at 194.

178. *See infra* text accompanying notes 125–128.

then, as a secondary safety valve. The major responsibility for regulating copyright's scope with respect to cases that pass the subject matter threshold is in the hands of a properly construed infringement test.[179] Only at the last stage, when copyrightable subject matter is used in a way that falls within copyright's proper scope, does fair use come into play as a doctrine of last resort that may nonetheless exempt certain uses on the basis of the case-specific analysis mandated by its factors.[180] Attempting to exercise the functions of subject matter rules (regulating the field's domain) or infringement analysis (primary regulation of copyright's scope) through the inadequate tool of fair use is the equivalent of turning a screw with a hammer, and hence a recipe for disaster. It carries with it the risk of conceptually confused and disjointed analysis that is likely not only to exacerbate the practical defects of fair use, but also generate incoherence throughout copyright law.[181]

The practical and conceptual dimensions of fair use, in combination, make it a particularly thin reed to lean on, especially in cases such as GenAI infringement that involve major technological and socio-economic disruption. The upheavals and high stakes associated with such cases attract intense pressures from stakeholders.[182] The fact- and law-intensive nature of the doctrine offers many opportunities for maneuvering and distinguishing precedent.[183] And the conceptual ill-fit of fair use for dealing with subject matter conflicts increases the risk of confusion by courts and incoherent or problematic decisions. At the end, using a hammer to do a screwdriver's job is a dangerous undertaking.

---

179. A different subject not discussed here is how some courts have departed from the properly construed infringement test and have eroded it in various ways. *See* Bracha, *Not De Minimis*, *supra* note 173, at 160.

180. *See* 17 U.S.C. § 107.

181. A perfect example of the complexity and incoherence generated by relegating a subject matter question to be handled by fair use is demonstrated by the recent Oracle Google litigation saga. The case involved the question of whether copying certain elements of computer code — the conventions of declaring code for Java language function libraries and the organizational structure of the libraries — was copyright infringement. The heart of the case was fundamentally a subject matter question of whether the kind of material copied is protectable by copyright. The district court ruled that it was not. Oracle Am., Inc. v. Google Inc., 872 F. Supp. 2d 974, 1002 (N.D. Cal. 2012). The Federal Circuit reversed the district court, nonetheless leaving the door open for fair use. Oracle Am., Inc. v. Google Inc., 750 F.3d 1339, 1381 (Fed. Cir. 2014). At the district court, the jury gave a verdict of fair use and the court rejected a motion for judgment as a matter of law, only to be reversed again by the Federal Circuit. Oracle Am., Inc. v. Google LLC, 886 F.3d 1179, 1211 (Fed. Cir. 2018). At the Supreme Court, the case was analyzed under fair use, resulting in a bitter disagreement between the majority that reversed the court of appeals and the dissent. Google LLC v. Oracle Am., Inc., 593 U.S. 1, 40–60 (2021). Thus, insisting on turning a subject matter case into a fair use one resulted in the case traveling up and down the courts with undulating results, a sharp divide at the highest court, and a final result that is broadly seen as providing little firm future guidance on the subject.

182. Lemley & Casey, *supra* note 119, at 763–70.

183. *Id.*

D. Doctrinal Application: Filtering

Correcting *Sega* by finding non-expressive copies to be non-infringing on subject matter grounds does not require much doctrinal innovation or reform. Nor does it require challenging the correct conclusion that training, or other non-expressive copies, fall within the statutory definition of "copies" and therefore constitute reproduction — they certainly do.[184] All it requires is applying the standard framework of copyright infringement analysis. At first, even if one is persuaded by the claim that non-expressive reproduction is outside copyright's domain, a puzzle seems to arise. The purpose of subject matter rules is to ensure that copyright protection extends only to informational subject matter that is within the field's domain. Yet, in the case of non-expressive copies, there is no claim that the copyrighted work consists of uncopyrightable subject matter. The denial that such copies infringe is not directed at the copyrighted work at all but at the alleged infringer's actions. How could this argument be about subject matter then? The puzzle is illusory. Copyright's subject matter rules are not a rigid logical structure that limits all analysis to a blinkered threshold examination of the copyrighted work. Instead, courts routinely incorporate subject matter analysis as a standard part of the infringement test by examining whether defendant took any protectable expressive elements.

Copyright's infringement test consists of two distinct elements: copying and improper appropriation.[185] The first is a purely factual question about the source of similarity between the works.[186] The second is a more normative inquiry that evaluates whether the copying is of an illicit character.[187] Improper appropriation is further divided into two inquiries.[188] First, courts examine which copied elements, if any, are protectable subject matter.[189] Second, courts assess whether there is

184. *See* 17 U.S.C. § 101 (defining "copies"); *see also* 17 U.S.C. § 106(1) (giving the owner the right to "reproduce the copyrighted work in copies").

185. Arnstein v. Porter, 154 F.2d 464, 468 (2d Cir. 1946).

186. *Id.*

187. *Id.* at 472–73.

188. Sturdza v. United Arab Emirates, 281 F.3d 1287, 1295–96 (D.C. Cir. 2002) (explaining the two steps of the "substantial similarity inquiry": identifying which copied elements are protectible and comparing substantial similarity of such elements); Harney v. Sony Pictures Television, Inc., 704 F.3d 173, 179 (1st Cir. 2013).

189. *See* Attia v. Soc'y of N.Y. Hosp., 201 F.3d 50, 54 (2d Cir. 1999) (observing that there may be "elements of a copyrighted work that are not protected even against intentional copying"); Palmer v. Braun, 287 F.3d 1325, 1330 (11th Cir. 2002) (explaining that copyright protection extends only to "original expression," and "does not extend to ideas, procedures, processes, or systems, regardless of their originality"); Kohus v. Mariol, 328 F.3d 848, 853 (6th Cir. 2003) ("[B]efore comparing similarities between two works a court should first identify and eliminate those elements that are unoriginal and therefore unprotected."); Johnson v. Gordon, 409 F.3d 12, 19 (1st Cir. 2005) (explaining that copying is not actionable if "an

sufficient similarity between the two works, based on copied protectable elements.[190] The first stage of the improper appropriation inquiry is where subject matter rules are incorporated into the infringement analysis. In this stage, courts filter out uncopyrightable copied elements before proceeding to examine similarity.[191] Its premise necessarily is that neither copying nor substantial similarity are sufficient in themselves to establish infringement. To infringe, there must be substantial similarity of copied *protectable* subject matter. And if no protectable elements were copied, no infringement is possible.

Finding non-expressive copies to be non-infringing on subject matter grounds is simply an application of this standard infringement test. Any physical reproduction of informational patterns that is completely detached from their expressive value should be filtered out at the first stage of the improper appropriation inquiry. Such reproduction involves no copyrightable subject matter and thus should be weeded out prior to evaluating similarity, in the exact same way that courts do with copying of functional elements, ideas, or scene-a-faire. In the case of completely non-expressive copies, following such filtering, no copied protectable subject matter remains. The outcome is the conclusion of no infringement.

One may object to this filtering infringement analysis by asking how the physical reproduction of the exact patterns of an expressive work involves no taking of copyrightable subject matter. But this would be simply falling back into the physicalist trap.[192] Reproduction of physical patterns as such is a mere physical fact that has no relevance for the subject matter of copyright law.[193] The physical fact of

---

impression [of substantial similarity] flows from similarities as to elements that are not themselves copyrightable").

190. *Sturdza*, 281 F.3d at 1296 ("Once unprotectible elements . . . are excluded, the next step of the inquiry involves determining whether the allegedly infringing work is 'substantially similar' to protectible elements of the artist's work.").

191. The origin of the term "filtering" comes from decisions regarding copyright protection for computer code. In that context, a special and particularly robust filtering framework has been applied by many courts. *See, e.g.*, Comput. Assocs. Intern., Inc. v. Altai, Inc., 982 F.2d 693, 707 (2d Cir. 1992) ("[W]e endorse[] a 'successive filtering method' for separating protectable expression from non-protectable material."). However, the analytical framework of a filtering step preceding the substantial similarity examination is broadly applied in all infringement cases. *See, e.g.*, Mattel, Inc. v. MGA Ent., Inc., 616 F.3d 904, 915 (9th Cir. 2010) (referring to the need to "filter out any unprotectable elements" when analyzing infringement claims with respect to sculpting plastic dolls); *Kohus*, 328 F.3d at 855 ("The essence of the first step [in analyzing improper appropriation] is to filter out the unoriginal, unprotectible elements . . . ."); Stromback v. New Line Cinema, 384 F.3d 283, 296 (6th Cir. 2006) (discussing "Filtering of Unprotected Elements"); Blehm v. Jacobs, 702 F.3d 1193, 1200 n.4 (10th Cir. 2012) ("We . . .filter out unprotected elements from the author's protected expression."); Rentmeester v. Nike, Inc. 883 F.3d 1111, 1118 (9th Cir. 2018) ("Before that comparison can be made, the court must 'filter out' the unprotectable elements of the plaintiff's work . . . .").

192. *See supra* Section IV.B.

193. *Id.*

reproduction is relevant only when it facilitates further access to expression which is copyright's proper subject matter. When that is not the case, the physically reproduced elements are non-expressive and therefore should be filtered out as part of the infringement analysis. And again, when the reproduction is completely non-expressive, no elements survive this stage, and the infringement test is not satisfied.

V. DOWNSTREAM: STYLE

A distinct unorthodox infringement argument raised against GenAI creation is that of copying of style.[194] It has elements of a traditional argument because it targets the output of GenAI, an information good that is itself a work whose expressive value is to be enjoyed by humans, unlike the non-expressive training copy. What makes the argument unorthodox is the level of abstraction at which it operates. The claim is not that any particular GenAI work is similar enough in the sum of its concrete details to any copyrighted work. Instead, what is being replicated in such cases is "style": more elusive and higher level of abstraction features of expressive works that, in this case, are associated with one particular creator.[195] As one dismayed artist aptly described the issue: "I can see my hand in it, but it is not my work."[196] The situation is dismaying to many because of the scale, speed, and ease involved. While borrowing or imitating style are hardly new phenomena in art and culture, GenAI brings it to a new level. Discerning existing patterns, across clusters of works, and implementing them to create new and different specific works is the core business of GenAI.[197] The result can be uncanny, especially when the output is generated in response to a prompt that requests for someone's specific style. But is such imitation of style infringing?

A. "Style" of a Single Work

First, we need to differentiate two different cases covered by the ambiguous concept of copying style. One case involves similarity to one specific copyrighted work, where the similarity is on a level of abstraction far-removed from verbatim copying. One could say that a particular painting is in the style of Roy Lichtenstein's famous work *In the*

194. *See, e.g.*, Complaint at 2, Andersen v. Stability AI Ltd., 23-cv-00201 (N.D. Cal. Aug. 12, 2024).

195. *See id.* (requesting relief for copyright infringement for "works generated by AI Image Products 'in the style' of a particular artist"); Andersen v. Stability AI, Ltd., 700 F. Supp. 3d 853, 860 (N.D. Cal. 2023) ("Plaintiffs allege that Stable Diffusion was 'trained' on plaintiffs' works of art to be able to produce Output Images 'in the style' of particular artists.").

196. Sylvie Douglis, *Artists vs. AI*, NPR (Jan. 30, 2023, 6:32 PM), https://www.npr.org/transcripts/1152653269 [https://perma.cc/N8GV-VJJ5].

197. *See supra* Part II.

Car. This would mean that the subsequent work, rather than being the same as the original with minor changes, copies high abstraction level themes and elements from the original. Perhaps the subsequent work depicts in an iconic comic book style a same-sex couple rather than a man and a woman, sitting side-by-side in the cockpit of a small aircraft with a different color scheme and angle compared to the original. A different case involves not a claim of similarities between two specific works, but one work incorporating multiple high-abstraction elements that together are characteristic of or are associated with a corpus of works by a particular creator. One might say, invoking this second sense, that a particular work is in the iconic Pop Art style of Roy Lichtenstein or in the style of Picasso's blue period.

The first variant can be dispensed with quickly. Saying that work A copies the style of work B is somewhat of a misnomer. The more accurate term is non-literal copying. Rather than claiming that work B is a literal copy of A or something close to it, the claim is that the similarities between the two specific works obtain on a higher abstraction and more diffused level. The question here is one of proper scope and the primary doctrinal tool for dealing with it is copyright's infringement test and its requirement of substantial similarity between the works.[198]

Two features of this analysis must be kept in mind. First, the standard for analysis under copyright's proper infringement test is whether the expression in defendant's work is similar enough to be a substitution for the original in its primary market.[199] This is in contrast with the recent tendency of some courts to replace this test with an anemic criterion under which all copying infringes unless it is only de minimis trivial taking from the original.[200] Second, as already discussed, subject matter rules are incorporated into the infringement test.[201] In a step sometimes described as "filtering," courts conceptually eliminate unprotectable elements prior to comparing the two works, to ensure that the similarity is of protectable expression.[202] Thus, certain common features between two works — which may be the basis of the copying of style claim — may be filtered out prior to conducting the substantial similarity analysis.[203] In short, copying of a single work's style should be analyzed under the standard infringement test. Under this test, the

198. *See* Arnstein v. Porter, 154 F.2d 464, 476 (2d Cir. 1946).

199. *See id.* at 473; *see also* Talha Syed & Oren Bracha, Copyright Rebooted 7 (Sept. 29, 2024) (unpublished manuscript) (on file with author).

200. *See* Bracha, *supra* note 173, at 158–69 (describing and criticizing the recent trend of some courts reducing the infringement test to an exception for de minimis copying).

201. *See supra* Section IV.D.

202. *See supra* note 189.

203. In this context, these features are likely to be: (1) certain high abstraction expressive elements that are treated as "ideas," *see* Nichols v. Universal Pictures Corp., 45 F.2d 119, 121–23 (2d Cir. 1930), and (2) stock elements within an expressive genre known as scènes à faire, *see* Abdin v. CBS Broad. Inc., 971 F.3d 57, 67 (2d Cir. 2020).

more abstract and less directly similar the copied elements are, the smaller the likelihood of infringement.[204]

B. "Style" of a Work Corpus

The argument that GenAI copies style is more likely the second variant of cases focused on a group of distinctive elements common to a corpus of works.[205] There are two distinct reasons why this argument fails.

The first relates to copyright's basic unit of protection. Copyright applies to "works of authorship."[206] A work is a discrete information good or an expressive package as created.[207] While courts are sometimes inappropriately lax in allowing plaintiffs to mix and match elements from expressive universes of different works, the proper and only unit of analysis in copyright law is the work.[208] Style is a contrived informational object constructed by combining elements from a corpus of multiple works. Copyright, however, applies to specific works, not to corpuses of works. As a result, substantial similarity to a corpus of works is an incoherent argument, one which is simply not recognized by copyright law.[209]

The second reason why the style argument fails, apart from the unit of analysis problem, is that style as a collection of distinctive and characteristic features that are spread across a corpus of works is not copyrightable subject matter under the domain side of the spillovers principle. A constructed information good such as the general style of Roy Lichtenstein or Jackson Pollock is considered too abstract to be protectable expression and, therefore, is not copyrightable subject

---

204. *See, e.g.*, Zalewski v. Cicero Builder Dev., Inc., 754 F.3d 95, 107 (2d Cir. 2014) ("Defendants' houses shared Plaintiff's general style, but took nothing from his original expression."); Ekern v. Sew/Fit Co., Inc., 622 F. Supp. 367, 369 (N.D. Ill. 1985) ("Copyright provides no protection to . . . writing styles."); Nesbitt v. Shultz, No. CV-00-0267, 2001 WL 34131675, at *7 n.4 (M.D. Pa. 2001) (holding that use of "flowery, over-drawn style, attempting to simulate a late-Victorian style of writing" is unprotectable because it is an idea); Douglas v. Osteen, 317 Fed. App'x. 97, 99 (3d Cir. 2009) ("[T]he use of a particular writing style or literary method is not protected by the Copyright Act.").

205. It appears that it is this meaning of style that the plaintiffs in *Andersen* have in mind. *See* Complaint at 1-2, Andersen v. Stability AI Ltd., No. 23-CV-00201 ("[T]he phrase 'in the style of,' refers to a work that others would accept as a work created by that artist whose 'style' was called upon.").

206. 17 U.S.C. § 102(a).

207. Oren Bracha & Talha Syed, Copyright's Atom 6 (Sept. 29, 2024) (unpublished manuscript) (on file with author).

208. *See id.* at 5. On the problem of the unit of analysis in copyright, see Margot E. Kaminski & Guy A. Rub, *Copyright's Framing Problem*, 64 UCLA L. REV. 1102, 1154–56 (2017).

209. *See* PATRY, *supra* note 90, at § 4:14 ("Style as a protectible element in U.S. copyright law is problematic [because] U.S. copyright law is 'work'-centric.").

matter.[210] Creators get to exclude others from their specific works, but their "style" is allowed to spill over to the public domain where everybody is free to learn from and reuse it. Allowing the borrowing of style is a crucial part of the spillovers principle's emphasis on permitting learning. It means that none may copy the works of Roy Lichtenstein or Jackson Pollock, but all are free to adopt their "style" to create different concrete works of their own. Doctrinally, copyright treats such elements as "abstract ideas" and denies them protection.[211]

However, we may still ask: should copyright law change to accommodate new circumstances? Arguably, GenAI creation introduces radically different circumstances that alter the relevant policy calculus. In the pre-AI days, one could argue, it was plausible to assume that limiting copyright protection to expression would suffice in most cases to cover the cost of creation and thus supply robust support for it. That was so even when other informational elements, such as style, were unprotected. Now, however, the remarkable ability of GenAI to replicate style has dramatically diminished the speed and cost of copying of such elements. The result is erosion of first-mover advantages with respect to them: GenAI catches up with creators before they get to extract the fruits of being the first innovators by developing their own unique expressive vocabulary.[212] As a result, copyright should recalculate its trajectory and extend the exclusion power to the hitherto unprotected elements of style.

Again, there are two reasons why this argument falls flat. First, the spillovers principle and the subject matter rules that embody it are not a shifting policy equilibrium, but a fundamental structural feature.[213] And there are good reasons not to try to optimize these rules by context. Once the principle of copyright's domain as strictly limited to expression is compromised, the shift is unlikely to stay restricted to the "troubling" GenAI context. Instead, the shift is likely to spread to other contexts, thereby undermining the basic structure of copyright and the principled policy balance embedded in it. The political economy of copyright exacerbates this concern. Supporting creativity via market-

---

210. *See* Samuelson, *supra* note 51, at 161 (observing that claims of similar style in GenAI output "seem weak because copyright law does not protect style as such").

211. *See* 17 U.S.C. § 102(a); Dave Grossman Designs, Inc. v. Bortin, 347 F. Supp. 1150, 1156–57 (N.D. Ill. 1972) ("Picasso may be entitled to a copyright on his portrait of three women painted in his Cubist motif. Any artist, however, may paint a picture of any subject in the Cubist motif, including a portrait of three women, and not violate Picasso's copyright so long as the second artist does not substantially copy Picasso's specific expression of his idea."); Williams v. 3DExport, No. 19-12240, 2020 WL 532418, at *3 (E.D. Mich. 2020) ("[E]ven if [the plaintiff] was the first to think up the anime, he could only have a protectible copyright interest in his specific expression of that idea.").

212. On first-mover or lead time advantages, see generally Stephen Breyer, *The Uneasy Case for Copyright: A Study of Copyright in Books, Photocopies, and Computer Programs*, 84 HARV. L. REV. 281, 299–300 (1970).

213. *See supra* text accompanying note 91.

based property rights comes with a built-in pressure for expansion by rent seekers.[214] Stable structural features such as the spillovers principle are exactly the field's partial remedy to this innate problem. Compromising these structural barriers in one context is likely to attract pressures that would destabilize them across copyright.

Second and more importantly, upon scrutiny, the claimed subject matter — style common to a corpus of works — is revealed to be a poor fit for the institutional form of copyright. The policy concern that copyright is designed to address is an appropriability-of-value problem related to a discrete information good and traceable to two features of such goods: non-excludability and the gap between creation and copying costs.[215] In short, the concern is that copying at a much lower cost than creating may undermine the ability of creators to recoup their cost of making specific works. Yet, style hardly fits the mold of a discrete information good, and, more importantly, the problem of creation/copying cost gaps plays only a minor role with respect to it. Corpus-wide style is a contrived information object, constructed through an exercise of conceptual abstraction: collecting elements from many different discrete works into one stitched together information good. Rather than a specific expressive work, style is more like a set of tools and building blocks, or a vocabulary. In this sense, style is not a discrete information good, but rather closer to the generative set of skills developed and possessed by a specific creator.

The character of style as closer to generative skills than to a concrete, consumable information good is reflected in the act of its appropriation. The reason why GenAI works that copy style seem so troubling is that they are significantly less costly to produce, even by comparison to new works by the creator whose style is being used: GenAI seems to be better (or at least more cost-efficient) than Warhol would be in producing a new Warhol. But the gap between the cost of developing a style and copying it plays only a minor role in this overall cost difference. The main reason why GenAI can outcompete the original creator in the market is that it is dramatically more cost-effective in creating something new by using the preexisting style elements. It is not so much that, unlike the human creator, it did not have to invest in developing the style.[216] In other words, the constitutive element of the copyright policy problem — free-riding attributable to not having to invest the development cost — plays only a trivial role. The real issue

214. *See* Bracha & Syed, *supra* note 207, at 9.

215. *See* Oren Bracha & Talha Syed, *Beyond the Incentive-Access Paradigm? Product Differentiation & Copyright Revisited*, 92 TEX. L. REV. 1841, 1849–50 (2014).

216. To be sure, GenAI creation also involves the considerable fixed costs of developing and operating the system itself. However, the overall cost efficiency, that explains the popularity of these systems, means that spread over a large enough number of users, these considerable fixed costs still do not prevent the cost of generating each new work from being considerably lower than those of human creation.

is that GenAI is simply more cost-effective in generating a new information good within a particular style, even compared to the creator whose style is being used.[217]

The real concern in appropriation-of-style cases is not the classic copyright policy problem, which focuses on specific information goods and creation/copying cost-gaps. Instead, it is a variant of a more general concern cutting across many fields in the wake of the rise of AI, namely the concern that AI, owing to its tremendous cost advantage, will take over production and displace humans. As I argue below, this indeed may be a genuine concern and one with unique implications in the field of creative expression, but copyright, which is focused on ameliorating problems rooted in the production/copying cost-gaps of specific works, is hardly the appropriate institutional tool for addressing this very different problem.[218]

Finally, a different concern that may be associated with copying of style is that of due social recognition to creators. In some cases, when style is copied with no acknowledgement of its source, a distinct harm may follow: the audience may falsely associate the GenAI work as originating with the human creator whose style is being used. The crucial point here is that this is a distinct harm with a distinct remedy. The pertinent concern is about recognition via express or implicit communication of accurate information about origin and credit. The appropriate remedy for such concerns is through a right for attribution, that is a right of persons to be accurately associated with works they created and not to be associated with works not created by them. Such an attribution right, including its negative aspect, is categorically different from a right to exclude. To a limited extent, a very narrowly applicable right of attribution already exists in copyright law.[219] To a larger although imperfect extent, attribution interests can be protected via trademark and unfair competition law.[220] There is certainly room for considering whether a more robust legal protection of the attribution interest is in order, including in light of GenAI appropriation of style. It remains the case, however, that such protection would be for accurate

---

217. To demonstrate this, consider a human creator and a GenAI who create a new comparable work developed within the style of the creator. Now ignore any past cost invested by the creator in developing the style elements (the informational element on which the GenAI is free riding). In many cases, the GenAI will still be much more cost-effective in generating the new work, even ignoring the element of cost spent in the past by the creator to generate the style. The upshot? The aspect that causes concern to many observers is not cost-gaps due to copying style elements, but cost-efficiency in new production that incorporates these elements.

218. *See infra* Part VI.

219. *See* 17 U.S.C. § 106A(a)(1).

220. *See* 15 U.S.C. § 1125(a) (Lanham Act section 43(a)). Existing law creates a complex relationship between copyright law on the one hand and unfair competition and trademark law on the other pertaining to designation of authorship in expressive works. *See, e.g.*, Dastar Corp. v. Twentieth Century Fox Film Corp., 539 U.S. 23, 33–37 (2003).

communication of meta-information about origin, not a right to exclude from using the style itself.

VI. COPYRIGHT'S LIMITS

In addition to preventing ownership of knowledge, copyright's subject matter principles play an additional, complementary role in scrutinizing unorthodox GenAI infringement arguments. This Part claims that such arguments are attempts to use broad copyright liability as a weapon for slowing down the rise of GenAI in markets for cultural expression. These attempts are driven by deep policy concerns about the possible effects of GenAI in the field of culture. This Part further argues that copyright is an inadequate institutional tool for addressing these concerns, no matter how genuine, and that subject matter rules operate to ensure that this area of the law is not burdened with tasks it is ill-equipped to perform. Unorthodox infringement arguments try to utilize copyright to solve broader social problems by designating as infringing activities that are not about the production and use of concrete expression. By insisting that copyright applies only to expression, subject matter rules place such claims outside of copyright's reach and prevent its application to policy problems for which it was not designed.

A. What Drives the Show

One may wonder what drives unorthodox, broad arguments for copyright infringement of either the upstream or downstream kind. Why not simply restrict claims to incidents of substantially similar generated output? In part, broad infringement claims are motivated by the prospect of imposing liability on entities along the GenAI supply chain with deeper pockets and strong control of the systems. But this is not the entire story. Unorthodox copyright infringement arguments appear to be translations into narrow legal forms of broader anxieties about the effects of GenAI on the creative realm. These anxieties are themselves variants of even broader apprehensions about the possible social effects of AI in general. While some of these concerns are speculative at this early stage, they may prove to be serious. However, the crucial question for our purposes is whether copyright is the right institutional space for dealing with the concerns.

Why might one be concerned about the effects of GenAI expressive generation in the realm of cultural creation? After all, the technology offers vast potential for high quality expressive production at low cost, more effective satisfaction of demand, wide and affordable availability, and even empowerment of human creativity via hybrid models

of creation.[221] The answer resides in three countervailing potential dangers, all of which are specific variants of more general concerns taken from an emerging litany of worries about the possible adverse effects of rapidly evolving AI on human society.[222]

The first concern is about the dissipation of sources of human employment and income in the creative industries. Indeed, the claim that GenAI's market-encroaching uses of works are infringing even if the GenAI's competing expressive product does not incorporate any protected expression is simply a version of this concern.[223] The fault of GenAI in this case is not that it learns to produce new competing works by extracting meta-knowledge from existing works. All creation does. Its fault is, rather, being so cost-effective in doing this that it threatens to outcompete many human producers out of the market. If GenAI can produce quality expression tailored to satisfying market demand at a fraction of the cost of human toil, wouldn't the result be displacing many professionals who make their living in the field?[224] This is, of course, a variant of the general anxiety about AI as a job killer.[225]

---

221. *See supra* text accompanying note 70.

222. This litany of concerns is growing long. It includes claims that are, while of possible tremendous consequence, sometimes hard to assess or even pin down exactly. *See, e.g.*, Yuval Harari, Tristan Harris & Aza Raskin, *You Can Have the Blue Pill or the Red Pill, and We're Out of Blue Pills*, N.Y. TIMES (Mar. 24, 2023), https://www.nytimes.com/2023/03/24/opinion/yuval-harari-ai-chatgpt.html [https://perma.cc/27G4-M6UW] ("A.I.'s new mastery of language means it can now hack and manipulate the operating system of civilization."); Future of Life Inst., *Pause Giant AI Experiments: An Open Letter* (Mar. 22, 2023), https://futureoflife.org/open-letter/pause-giant-ai-experiments [https://perma.cc/C3U9-N3VK] ("Should we develop nonhuman minds that might eventually outnumber, outsmart, obsolete and replace us? Should we risk loss of control of our civilization?"). Eliezer Yudkowsky, *Pausing AI Developments Isn't Enough. We Need to Shut it All Down*, TIME (Mar. 29, 2023, 6:01 PM), https://time.com/6266923/ai-eliezer-yudkowsky-open-letter-not-enough [https://perma.cc/7SUU-85TL] (predicting that "the most likely result of building a superhumanly smart AI, under anything remotely like the current circumstances, is that literally everyone on Earth will die").

223. *See* Sobel, *supra* note 129, at 76–79.

224. *See* Samuelson, *supra* note 51, at 159 ("Generative AI seems poised to have substantial impacts on the careers of professional writers and artists."); Mandalit del Barco, *Striking Hollywood Scribes Ponder AI in the Writer's Room*, NPR (May 18, 2023, 8:52 PM), https://www.npr.org/2023/05/18/1176876301/striking-hollywood-writers-contemplate-ai [https://perma.cc/2FL3-TX2T]. For an example that directly connects concerns about displacing human creators to struggles, including legal ones, against GenAI creation, see Vanessa Thorpe, *'ChatGPT Said I Did Not Exist': How Artists and Writers Are Fighting Back Against AI*, THE GUARDIAN (Mar. 18, 2023, 12:00 PM), https://www.theguardian.com/technology/2023/mar/18/chatgpt-said-i-did-not-exist-how-artists-and-writers-are-fighting-back-against-ai [https://perma.cc/X4ZP-UVGJ].

225. *See, e.g.*, Steven Greenhouse, *US Experts Warn AI Likely to Kill Off Jobs — And Widen Wealth Inequality*, THE GUARDIAN (Feb. 8, 2023, 2:00 AM), https://www.theguardian.com/technology/2023/feb/08/ai-chatgpt-jobs-economy-inequality [https://perma.cc/6WQY-KGQX]; Annie Lowrey, *How ChatGPT Will Destabilize White-Collar Work*, THE ATLANTIC (Jan. 20, 2023), https://www.theatlantic.com/ideas/archive/2023/01/chatgpt-ai-economy-automation-jobs/672767/ [https://perma.cc/6ASN-C7NZ].

While some preliminary studies suggest that the concern may be founded in some areas, it is extremely hard to assess in the field of expression.[226] Some high-flying human creators, especially creators of auratic works — where much of the work's value is attributable to its embodiment in a unique physical copy — are likely to maintain demand for their higher cost product. [227] Indeed, in an age of machine production rather than just reproduction, the value of auratic works made by human hands may increase.[228] Furthermore, GenAI expression may be used as a tool rather than a substitute for human creativity, especially through various hybrid models of creation where machines do not completely take over expressive agency.[229] It is hard and dangerous to predict what would be the bottom line of the countervailing effects of GenAI on human involvement in supplying market demand for expression. Still, what if the dire predictions come to pass and GenAI ends up being a killer of creative jobs, at least in some sectors? What if legions of positions of graphic designers, script writers, video editors and sound engineers are eliminated with no substitutes?

The other two distinct concerns are both about the shrinking space for human creativity.[230] Like the first, both of these additional concerns are variants in the cultural sphere of a more general anxiety about AI whose focus is the effects on human lives, capacities, and potentials of outsourcing to machines activities traditionally entrusted to human intelligence. These two concerns are distinct from the previous concern

---

226. *See, e.g.*, Tyna Eloundou, Sam Manning, Pamela Mishkin & Daniel Rock, GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models 3 (Aug. 21, 2023) (unpublished manuscript) (on file with arXiv), https://arxiv.org/abs/2303.10130 [https://perma.cc/Q9VS-WJDX] (estimating that "up to 49% of workers could have half or more of their tasks exposed to" Large Language Models AI); Edward W. Felten, Manav Raj & Robert Seamans, How Will Language Modelers Like ChatGPT Affect Occupations and Industries? 2 (Mar. 18, 2023) (unpublished manuscript) (on file with arXiv), https://arxiv.org/abs/2303.01157 [https://perma.cc/897V-GRSY] (developing and applying a methodology "to identify which occupations, industries and geographies are most exposed to AI").

227. The term auratic works is based on Walter Benjamin's concept of aura. *See* Walter Benjamin, *The Work of Art in the Age of Mechanical Reproduction*, *in* ILLUMINATIONS 181, 184 (Hannah Arendt ed., 1969). The term refers to expressive works that according to Benjamin have a "quality of its presence" derived from the authenticity and authority of a unique physical object in which the work is embedded. *Id.* Benjamin famously argued that in the age of mechanical reproduction, where the existence of one unique authentic object embodying the work disappears, aura withers. *Id.*

228. *See* Stefan Bechtold & Christopher Jon Sprigman, *Intellectual Property and the Manufacture of Aura*, at 4 (N.Y.U. Sch. L., Pub. L. Research Paper No. 22-09, 2022), https://ssrn.com/abstract=4002717 [https://perma.cc/7Y87-GWUG] (arguing that "auratic experience is engineered through a combination of reproduction techniques, social norms, community building, and interlocking business and legal strategies" even in the age of digital reproduction).

229. *See supra* text accompanying note 70.

230. *See* Daniel J. Gervais, *The Human Cause*, *in* RSCH. HANDBOOK ON INTELL. PROP. AND A.I. 22, 35 (Ryan Abbott ed., 2022) ("Ultimately, the risk of replacing humans in the act of creation, perhaps our noblest quest, is the principal consideration.").

because they are not about access to the economic fruits of creativity but about the value of human creativity itself. Harari, Harris, and Raskin, for example, ask: "What would it mean for humans to live in a world where a large percentage of stories, melodies, images, laws, policies and tools are shaped by nonhuman intelligence[?]" and answer that "A.I. could rapidly eat the whole of human culture."[231] Their formula is vague and may sweep in many different fears.[232] Internal to the field of expression, however, one could identify two separate dangers of entrusting the development of large chunks of culture to the hands of machines: one is about the intrinsic value of human creativity and the other pertains to the sources of pathbreaking innovation.

As to the second concern whose focus is the intrinsic value of creativity, consider one question posed by an open letter recently published on the subject of regulating AI: "Should we automate away all the jobs, including the fulfilling ones?"[233] The point is that expressive creativity has its own inherent value to those who engage in it, well beyond satisfying market demand. Opportunities to access the benefits of this inherent value would be lost if creativity is transferred from humans to machines.[234] The diehard market economist might be puzzled at this point. She might ask: If expressive activity has value to creators that outweighs its higher cost when practiced by humans, why would rational creators not swallow the cost and undertake such activity? And if the intrinsic value does not outweigh the cost, is it not obvious that it is better for the activity not to happen? The failure of such questions is reducing all value to market value. Dismissing them requires a normative framework other than maximizing market efficiency.[235]

Several such frameworks are, in fact, available. One focuses on distributive equity.[236] In our context, it would raise concerns about broad and fair distribution of access to expressive activity.[237] Other frameworks are centered either on an affirmative account of the good

---

231. *See* Harari et al., *supra* note 222.

232. Harari et al.'s own concern here seems to be focused on Matrix-like scenarios of manipulation by artificial intelligence of constitutive human "culture" for nefarious purposes or of losing human control altogether to intelligent machine manipulation. *See* THE MATRIX (Warner Bros. Pictures 1999).

233. Future of Life Inst., *supra* note 222.

234. *See, e.g.*, Thorpe *supra* note 224 (statement of Susie Alegre) ("Do we really need to find other ways to do things that people enjoy doing anyway? Things that give us a sense of achievement, like writing a poem?").

235. *See generally* Bracha & Syed, *supra* note 115 (mapping democratic theories of copyright and the ways in which they differ from efficiency).

236. *See id.* at 287–313; Amy Kapczynski, *The Cost of Price: Why and How to Get Beyond Intellectual Property Internalism*, 59 UCLA L. REV. 970, 993–1006 (2012).

237. This is especially so given the fact that market-based access is based on the criterion of ability to pay. *See* Kapczynski, *supra* note 236, at 996; *see generally* Molly Shaffer & Van Houweling, *Distributive Values in Copyright*, 83 TEX. L. REV. 1535 (2005) (exploring copyright's impact on equitable opportunities for expressive activity).

human life[238] or on higher-order values that validate and give significance to individual market preferences.[239] The first translates here into arguments about the intrinsic value of expressive activity as a variant of meaningful activity which is essential for human self-realization.[240] The second highlights the importance of a diverse and robust sphere of human expression for individual, collective, and cultural self-determination.[241] The upshot of all of this is that loss of opportunities for realizing the intrinsic value of creativity cannot be reduced into a cost-benefit calculus measured by market value.

This concern is shored-up by the fact that the intrinsic value of creativity involves an acquired capacity: to enjoy it one needs to develop the taste for it through the practice of exercising the relevant human capacities.[242] The problem thus has an endogeneity aspect: not only is the intrinsic value of creativity unreducible to subjective preferences, the preferences and subjective value may not develop in the first place without robust opportunities to experience the activity.[243]

The third concern is about dwindling sources of deep innovation. GenAI's business is low-cost generation of new expressive goods within established patterns.[244] The troubling question is: if much of the human creative sphere is eliminated by being offshored to machines, where would pattern-breaking, disruptive innovations come from? GenAI can churn out low-cost impressive new pictorial, musical, or textual works within established patterns. But where would the next Impressionism, Pop-Art, hip-hop, new wave cinema, or Dada come from?[245] Once more, a market-oriented person could answer: if there is demand for it and machines cannot provide it, someone will. The nature of disruptive expressive innovation, however, is exactly that it breaks

---

238. *See* Bracha & Syed, *supra* note 115, at 256–58.

239. *See id.* at 251–56.

240. *See, e.g.*, William W. Fisher III, *Reconstructing the Fair Use Doctrine*, 101 HARV. L. REV. 1659, 1744–66 (1988); *see generally* MADHAVI SUNDER, FROM GOODS TO A GOOD LIFE: INTELLECTUAL PROPERTY AND GLOBAL JUSTICE (2012).

241. *See* Bracha & Syed, *supra* note 115, at 251–56.

242. *See id.* at 279–80. The argument follows a similar structure to Dworkin's argument in support of a liberal state's subsidization of the arts based on the claim that "it is better for people to have complexity and depth in the forms of life open to them." *See* RONALD DWORKIN, *Can a Liberal State Support Art?*, *in* A MATTER OF PRINCIPLE 221, 229 (1985).

243. The idea of institutions designed not merely to satisfy preferences but to allow people to develop preferences and capacities is traceable to John Stuart Mill. *See* Michael S. McPherson, *Mill's Moral Theory and The Problem of Preference Change*, 92 ETHICS 252, 255 (1982); JOHN STUART MILL, AUTOBIOGRAPHY 97–98 (Oxford Univ. Press 2018) (1873) (observing that the design of institutions is more than one of "material interests" and "ought to be decided mainly by the consideration, what great improvement in life and culture stands next in order for the people concerned, as the condition of their further progress, and what institutions are most likely to promote that").

244. *See supra* Part II.

245. *See* Gervais, *supra* note 230, at 35 (observing that "[t]he risk is that there will be more of the same, or worse").

from existing established patterns and existing tastes.[246] Therefore, even if one accepted the questionable theory that demand attracts creativity rather than assume more intrinsic motivation, it is the disruptive character of the creativity in question — the fact that it diverges from preexisting demand patterns — that undermines the claimed flow from demand to production.

More importantly, creative innovation does not simply happen out of thin air. Various social and environmental conditions shape the degree to which the climate is hospitable for innovation.[247] Key among those is the existence of a robust social sphere of creativity where people engage in the practices that develop the relevant skills, tastes, and inclinations.[248] The concern is that a dwindling of a robust market-supported human sphere of creativity would also impoverish the social conditions, as well as human potential and capacities that are the soil from which disruptive creativity sprouts.

B. Where Copyright Runs Out

The specter of GenAI taking over much of the space for human creativity does raise, then, serious concerns about losing sources of livelihood, the intrinsic value of creative activity, and the social value of disruptive innovation. The remaining question is whether copyright is an adequate institutional tool for addressing these concerns.[249] Examining the policy problem around which copyright was designed and the institutional tools with which it is equipped reveals that turning to it for solutions would be dubious, even if the feared maladies are real.[250]

The main copyright strategy, presumably designed to face the threats of GenAI, is hampering GenAI entry into markets for cultural expression by erecting high proprietary walls. These walls are founded

---

246. *See* NEIL WEINSTOCK NETANEL, COPYRIGHT'S PARADOX 40 (2008) (discussing "oppositional expression"); Bracha & Syed, *supra* note 115, at 269–70 (discussing "heterodox works"). *See generally* JOSEPH A. SCHUMPETER, CAPITALISM, SOCIALISM AND DEMOCRACY 82–83 (1994) (discussing his concept of "creative destruction" in innovative economic activity).

247. *See* Xavier Kastañer & Lorenzo Campos, *The Determinants of Artistic Innovation: Bringing in the Role of Organizations*, 26 J. CULTURAL ECON. 29, 34–40 (2002) (reviewing the literature on the social determinants of artistic innovation).

248. *See* Neil Weinstock Netanel, *Copyright and a Democratic Civil Society*, 106 YALE L.J. 283, 343–44 (1996) (discussing the analogous context of participation in the public sphere as necessary for individuals developing the expressive skills and habits necessary for agency as a citizen).

249. *See generally* Blake E. Reid, *What Copyright Can't Do*, 52 PEPP. L. REV. (forthcoming 2025) (manuscript at 1) (exploring the "practical limits in the structure of contemporary copyright law and doctrine that constrain copyright's capabilities for solving policy problems beyond copyright's usual ambit").

250. For a similar argument, see Ard, *supra* note 123, at 359 ("Copyright provides a remarkably limited toolkit for dealing with the problems of AI — even the problems it poses for artists.").

on broad, unconventional infringement arguments that target either the training process or the generated output.[251] One problem with this strategy is that its effects are unlikely to remain confined to the GenAI context. Once the spillovers principle is compromised and learning metaknowledge and other abstract elements become protectable, this shift is likely to bleed elsewhere and undermine the balance built into the structure of copyright. After all, if taking of style and extracting metaknowledge in learning how to create independent expression can infringe in the GenAI context, why not elsewhere?

More importantly, the copyright strategy is hardly appealing on its own terms. Throwing a wrench into the wheels of GenAI development in the form of newly expanded, taxing versions of copyright liability is far from an attractive solution. Just as breaking up the machines was hardly an adequate solution for the social ills brought about by early industrialization, smashing creative machines via copyright appears as an unattractive form of neo-Luddism.[252] Instead of smashing the machines, the goal should be creating a legal framework that allows enjoying the benefits of the technology in supplying market demand and enabling human creativity, while holding its potential dangers at bay. That is not a job fit for copyright. Copyright's institutional tools are crude in this context. All they can achieve is market barriers to entry of GenAI through prohibitive costs, and perhaps limited market transfers to some human creators, in cases where licensing does take place. Attempting to balance the general social costs and benefits of GenAI by calibrating copyright entitlements seems an intractable or even fantastic task.

Copyright is a poor fit for the task because it was designed for other purposes. At the bottom of copyright's inaptness for meeting the broad challenges of GenAI lies a deep incongruity between its basic logic and the relevant social concerns. There are three interlocking aspects to this incongruity: the policy problem copyright was designed to address, the information good to which it applies, and its institutional form as a market-based mechanism.

First, the policy problem that animates copyright is fundamentally different from those that are created by GenAI. Copyright is designed

---

251. *See supra* Section IV.A, Part V.

252. The Luddites, named after Ned Ludd, formed a movement of English textile workers who opposed cost-saving machinery in the early nineteenth century. *See generally* BRIAN BAILEY, THE LUDDITE REBELLION (1998). They are known today for their opposition to industrial machines and the social upheavals of industrialization, sabotage and machine breaking, and the ultimate violent oppression of the movement. *See* E.J. Hobsbawm, *The Machine Breakers*, 1 PAST & PRESENT 57, 58 (1952). The term has come to refer more generally to opposition to industrialization, automation, and new technologies. *Id.* As Eric Hobsbawm has argued, machine-breaking was a multifaceted phenomenon that was not limited to principled opposition to machines. *See id.* at 58–59 (distinguishing between machine breaking as a form of effective "collective bargaining by riot" and a less dominant version based on hostility to machines).

around a specific public policy problem: the inability of creators to appropriate sufficient value from their creation due to the gap between creation and copying cost and the difficulty of excluding others.[253] Copyright ameliorates this problem by conferring a limited right to exclude that is designed to increase appropriability while keeping the resultant cost on access in check.[254] Some cases of GenAI production of expression — the paradigmatic example is generated output substantially similar to copyrighted works — fit this mold. But the broader concerns that drive unorthodox infringement arguments do not. These concerns are not about inability to recoup the cost of creation due to low-cost copying. They are about the social implications of machines that can outcompete human creators in satisfying market demand through cost-effective generation of new works.[255] It is not surprising that the tools of copyright that are designed to address one kind of social policy problem do not fit when they are expected to solve a completely different one. As recently demonstrated in other digital policy contexts, when copyright is wielded to combat troubling problems that are beyond its core concerns and institutional tools, the results are often poor.[256]

Second and on a deeper level, there are two distinct information goods involved. The information good that is the focus of copyright is completely different from the information that drives the social dynamics at the heart of the GenAI policy concerns. Copyright law and policy is focused on the value of a discrete information good: the expressive value of specific works. The appropriation policy problem that drives copyright pertains to the dynamic of producing and consuming these discrete information goods. The GenAI policy concerns are driven by the generation and use of a different kind of information: social information about expressive goods.[257] This social information is meta-information: it is not the information (whether content or expressive form) contained in specific works, but rather information about regularities and relations in the informational patterns of such works. It is *social* information because its value consists in aggregating patterns common to many individual expressive works. Meta-information on the patterns of one or even very few works would be useless for the GenAI process. The value of this information is in its broad social-

253. *See* Bracha & Syed, *supra* note 115, at 237–38.

254. *See id.* at 238–40.

255. *See supra* Section VI.A.

256. *See* Neil Netanel, *The EU Press Publishers' Right Is Inapt and Off-Target*, 46 COLUM. J.L. & ARTS 301, 301 (2023) (explaining that "any copyright the press publisher has in news articles are likely to be ineffective in addressing newsrooms' dramatic loss of revenue in recent decades"); Neil Weinstock Netanel, *Mandating Digital Platform Support for Quality Journalism*, 34 HARV. J.L. & TECH. 473, 496–98 (2021).

257. *See* Salomé Viljoen, *A Relational Theory of Data Governance*, 131 YALE L.J. 573, 603–13 (2021) (discussing social data including its "horizontal" relational aspect).

relational orientation.[258] To be sure, the social information is leveraged to produce new concrete informational works with discrete value at the generation stage, but the driving force of the process is the social and relational aspects of aggregated information.

This is the deeper source of the mismatch. Copyright regulates the production/use dynamics of discrete information goods, while the broad GenAI policy concerns are focused on a process powered by aggregated social meta-information about these goods. The process of producing and using this social information may have good aspects (cost-effective satisfaction of market demand and empowering creativity via hybrid human-machine models) and bad aspects. However, attempting to ameliorate the bad aspects through a legal regime made for the regulation of a different kind of information good is bound to misfire.[259]

Third, copyright's market-focus stands in deep tension with the extra-market basis of the policy concerns about GenAI. Copyright is a market-based mechanism.[260] As such, it has familiar relative strengths and weaknesses with respect to its core mission.[261] The GenAI policy concerns, however, go beyond market-based logic. Each concern challenges or questions in different ways the outcomes dictated by pure market efficiency, namely the dominance of the most cost-effective means for stratifying demand.[262] If maximizing social wealth as measured by market value were the overarching concern, there would be no reason to be worried about cost-effective production resulting in loss of

---

258. An analogous context is aggregated social information about individual behavior. *See id.* at 611–12 ("In a typical data flow, any one individual's data is essentially meaningless, and the marginal cost of any one individual defecting from collection is very low.").

259. What is at work here is the same mismatch that characterizes trying to regulate misuses of aggregate social information about people and their behavioral patterns by using the institutional tools of personal privacy that are focused on individual rights against misusing individual, personal information. *See id.* at 617 ("Both [existing proposals for privacy law reforms] attempt to reduce legal interests in information to individualist claims subject to individualist remedies that are structurally incapable of representing the horizontal, population-level interests of data production. This in turn allows significant forms of social informational harm to go unaddressed and may foreclose socially valuable forms of data production.").

260. *See generally* Harold Demsetz, *Information and Efficiency: Another Viewpoint*, 12 J.L. & ECON. 1 (1969).

261. With respect to competing institutional tools in intellectual property, see Kenneth Arrow, *Economic Welfare and the Allocation of Resources for Invention*, *in* NAT'L BUREAU OF ECON. RESEARCH, THE RATE AND DIRECTION OF INVENTIVE ACTIVITY 609 (R.R. Nelson, ed., 1962); Richard R. Nelson, *Uncertainty, Learning, and the Economics of Parallel Research and Development*, 43 REV. OF ECON. & STAT. 351 (1961); Yochai Benkler, *Intellectual Property and the Organization of Information Production*, 22 INT'L REV. L. & ECON. 81 (2002).

262. *See supra* Section VI.A.

jobs, opportunities for creative activity, or potential for disruptive innovation.[263]

This discontinuity between copyright's market focus and the extra-market drive of the policy concerns limits its effectiveness. One can hardly track extra-market policies by simply mobilizing the market via property rights or limiting entry to it.[264] Instead the market should be embedded.[265] That is to say, the legal framework should facilitate reaping the fruits of machine creation in satisfying demand and empowering human creativity, while simultaneously embedding this activity within mechanisms that serve extra-market interests: ensuring equitable sources of basic income, fostering opportunities for human creativity for its intrinsic value, and cultivating the social conditions from which disruptive innovation can arise. How exactly to achieve this is beyond the scope of this Article. One can safely predict, however, that the appropriate vehicle would not be copyright, but rather an arsenal of different institutional tools including the regulation or taxation of GenAI technology on the one hand and affirmative support and cultivation of opportunities for human creativity on the other.

If unorthodox infringement arguments are attempts to use copyright to solve AI-related policy problems that copyright is ill-suited to handle, how can we ensure that copyright is not burdened with tasks that are beyond its ken? This is exactly the function of copyright subject matter rules and their application to reject unorthodox infringement arguments both upstream and downstream, as it was described in this Article.[266] These rules restrict copyright to the kind of policy problems it is equipped to handle and channel away from it those it is not. Subject matter rules carry out this function by making sure that copyright is only applied to the kind of informational subject matter it was designed and equipped to handle — namely expression — while all other issues,

---

263. The only possible concern would be about distribution grounded in either fairness or more likely in considerations of diminishing marginal utility. The standard recommendation for ameliorating such concerns would be post-market-activity redistribution via taxation. *See* Louis Kaplow & Steven Shavell, *Why the Legal System is Less Efficient than the Income Tax in Redistributing Income*, 23 J. LEGAL STUD. 667, 667–68 (1994).

264. The claim that as a market-based mechanism copyright has certain built-in institutional limits owing to constraints stemming from the basic logic of markets is one that might raise eyebrows among those familiar with realist or Critical Legal Studies critiques of any "logic" to markets or other institutions given their "legal indeterminacy." For what is probably the most famous version of this critique, see ROBERTO MANGABERIA UNGER, PLASTICITY INTO POWER, VOL. 3 OF POLITICS: A WORK IN CONSTRUCTIVE SOCIAL THEORY 69 (1987). *See also* Samuel Moyn, *Thomas Piketty and the Future of Legal Scholarship*, 128 HARV. L. REV. 49, 55 (2014) (echoing the claim that "there is no such thing as capitalism"). For a reply to such arguments from legal indeterminacy, see Talha Syed, Legal Realism and CLS from an LPE Perspective 54–68 (Nov. 14, 2023), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4601701 [https://perma.cc/QB3E-6BBU].

265. *See* Talha Syed, Embedding Innovation 6 (Sept. 29, 2024) (unpublished manuscript) (on file with the author).

266. *See supra* Sections IV.C, V.B.

even when genuine policy problems are involved, are kept out of its domain. Thus, preventing unorthodox GenAI infringement arguments from assigning to copyright social policy problems it cannot handle is exactly the work done by subject matter rules when they are properly applied in this area.

VII. CONCLUSION

This Article has argued that resorting to copyright's fundamental principles and their purpose is necessary for facing the challenges posed by GenAI infringement. Specifically, copyright's subject matter rules that regulate the field's domain under the spillovers principle have an important role to play. Subject matter rules dispose summarily of the most ambitious, unorthodox GenAI infringement claims both upstream and downstream. Upstream, non-expressive training copies are a variant of an activity that has always been privileged by copyright law: the extraction of meta-information for the purpose of learning. Despite the incident of physical reproduction, the activity does not trigger copyright's sine qua non — the use and enjoyment of expressive value — and thus is outside copyright's domain. On the downstream side, copying of style arguments equally fail on subject matter grounds. The appropriation of style claim attempts to construct a cross-work information good that is not recognized by copyright and then extend protection to unprotectable high-abstraction elements of that good.

Proper application of subject matter rules also helps avoid the danger of charging copyright with work it is ill-suited to perform. Attempts to extend novel and broad forms of copyright liability are fueled by fears of various deleterious effects of AI in the realm of culture. While some of these concerns may prove justified, expanded copyright liability targeting GenAI hardly seems to be an adequate response. Given the policy problem around which it was designed, the information good that is its focus, and the institutional tools at its disposal, copyright is a very poor fit for solving the new problems. Indeed, should the prediction of GenAI dominance in markets for expression materialize, the centrality of copyright as a policy tool for supporting human creativity might be somewhat diminished. After all, market-based internalization via price premiums obtained by legal exclusion — which is a very long way of saying "copyright" — has only existed as the primary social means for supporting the production of costly human expression for the last 300 years or so. Copyright has served us well, but its anchor in the market has always been a somewhat uneasy fit for the values pertaining

to expression.[267] To the extent that the rise of GenAI causes the widening of this gap between copyright's market basis and the values we seek to uphold in the cultural sphere, perhaps it will also mark a relative shift to other mechanisms for supporting human creativity. Time will tell.

267. *See* Neil Weinstock Netanel, *Market Hierarchy and Copyright in Our System of Free Expression*, 53 VAND. L. REV. 1879, 1899 (2000) (discussing how copyright contributes to "speech hierarchy").