COPYRIGHT'S LATENT SPACE: GENERATIVE AI AND THE LIMITS OF FAIR USE

BJ Ard†

Generative AI poses deep questions for copyright law because it defies the assumptions behind existing legal frameworks. The tension surfaces most clearly in debates over fair use, where established tests falter in the face of generative systems' distinctive features. This Article takes up the fair-use question to expose copyright's limitations as well as its latent commitments, particularly its allowances for the exploitation of non-authorial value.

Fair use's transformative use paradigm, which compares the purpose of the use with that of the original work, faces difficulty evaluating copying during the training of AI models. Close examination of the technology—from training through the operation of completed systems—reveals that the purpose of copying may be contingent because a model's capabilities and ultimate uses are indeterminate at the time of training. This hurdle can be sidestepped by recognizing that purpose serves as a proxy for determining whether the use intrudes on markets rightly belonging to the copyright owner. However, this raises the question of which markets those are.

Answering the market question requires delving into copyright's latent space—the unarticulated principles and

† Associate Professor of Law, University of Wisconsin Law School, and Affiliate Fellow, Yale Information Society Project. I thank Oren Bracha, Robert Brauneis, Carys Craig, R. Feder Cooper, Rebecca Crootof, Deven Desai, Justin Hughes, Mark Lemley, Glynn Lunney, Michael Murray, David Nimmer, Jacob Noti-Victor, Blake Reid, Matthew Sag, Frederic Sala, Benjamin Sobel, Madhavi Sunder, Charlotte Tschider, Molly Shaffer Van Houweling, Christopher Yoo, and Peter Yu for feedback on this project, along with participants at the Fifth Annual Art Law Works-in-Progress Colloquium, 2024 Copyright Scholarship Roundtable, UW Integrating Robots into the Future of Work Colloquium, 2024 Legal Scholars Roundtable on AI, Texas A&M School of Law Transformation in IP and Technology Law Symposium, 2024 Works-in-Progress for IP Colloquium, and faculty workshops at UGA and UW law schools. I also thank Kate Bishop, Peter Feider, John Lavanga, and Rosemary Patton for excellent research assistance. Research support was provided by the Office of the Vice Chancellor for Research and Graduate Education at UW with funding from the Wisconsin Alumni Research Foundation.

commitments embedded in its jurisprudence. This Article identifies a dividing line between market value that stems from an author's creative choices and market value that does not, with courts permitting users to tap into the latter even to the copyright owner's detriment. The reoriented test would ask whether a user exploits non-authorial value like that which stems from facts, tropes, and third-party investment versus the authorial value arising from an artist's creative decisions. The precise line remains to be hashed out—courts have historically drawn the line differently across creative fields to balance copyright's competing objectives in specific contexts.

The fair-use question also reveals deeper structural limitations of the copyright regime. Concretely, the argument that copyright's pro-artist policies compel denial of fair use misses that AI systems trained on licensed works may still displace human creators. The lack of unauthorized use takes the problem outside copyright's domain. The core problem is not the duplication of specific works, but the ability to produce comparable works cheaply and quickly. The challenge cannot be resolved through the mere extension or denial of fair use. Instead, it demands we put copyright in dialogue with other regimes for promoting the arts, blunting the misuse of these tools, and confronting the technology's capacity to consolidate power.

Introduction

Generative AI has been embroiled in controversy since its public debut in 2022. Although services like ChatGPT have entered widespread use,[1] vocal segments of the public have painted the creation of these systems as theft.[2] Art systems have trained on hundreds of millions,[3] if not billions, of images scraped from the internet,[4] and language models have trained

---

1 *See* Colleen McClain, *Americans' Use of ChatGPT Is Ticking Up, But Few Trust Its Election Information*, Pew Rsch. Ctr. (Mar. 26, 2024), https://www.pewresearch.org/short-reads/2024/03/26/americans-use-of-chatgpt-is-ticking-up-but-few-trust-its-election-information [https://perma.cc/V7ZT-A8NE].

2 "AI-art generators are trained on enormous datasets, containing millions upon millions of copyrighted images, harvested without their creator's knowledge, let alone compensation or consent. This is effectively the greatest art heist in history." *Restrict AI Illustration from Publishing: An Open Letter*, Ctr. for Artistic Inquiry & Reporting (May 2, 2023), https://artisticinquiry.org/AI-Open-Letter [https://perma.cc/2ZW8-NQN7].

3 Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu & Mark Chen, *Hierarchical Text-Conditional Image Generation with CLIP Latents*, arXiv 23 (Apr. 13, 2022), https://arxiv.org/pdf/2204.06125 [https://perma.cc/A977-BZJ6] [hereinafter DALL-E 2 Paper] (training on 650 million image-text pairs).

4 *See* Benj Edwards, *Meta's New AI Image Generator Was Trained on 1.1 Billion Instagram and Facebook Photos*, Ars Technica (Dec. 6, 2023), https://arstechnica.com/

on corpuses of hundreds of billions of words taken from books, web pages, and other texts.[5] Creative and knowledge workers of every stripe, lawyers included, fear for their professions' futures because generative AI produces cheap, passable substitutes for human creativity.[6] Meanwhile, the FTC has detailed a range of AI concerns ranging from deepfakes, fraud, and privacy violations[7] to the entrenchment of monopolistic practices among major tech companies and platforms.[8]

Not surprisingly, copyright is central to the ongoing debate. One of the most contentious questions is whether the literal copying that occurs during training qualifies as fair use.[9]

---

information-technology/2023/12/metas-new-ai-image-generator-was-trained-on-1-1-billion-instagram-and-facebook-photos [https://perma.cc/U77P-UV9M].

5 *See* Complaint, P.M. v. OpenAI LP at ¶ 146, No. 3:23-cv-03199, 2023 WL 4335507 (N.D. Cal. June 28, 2023) (alleging OpenAI "systematically scraped 300 billion words").

6 *See* Tyna Eloundou, Sam Manning, Pamela Mishkin & Daniel Rock, *GPTs Are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models*, ARXIV (Aug. 22, 2023), https://arxiv.org/pdf/2303.10130 [https://perma.cc/P88B-XBXS] (predicting impacts for 80% of the U.S. workforce); Copyright All., Comment Letter on U.S. Copyright Office Notice of Inquiry on Artificial Intelligence and Copyright 95 (Oct. 30, 2023), https://www.regulations.gov/comment/COLC-2023-0006-8935 [https://perma.cc/Q58B-RC3L] (detailing threats to artists).

7 *See* Press Release, Federal Trade Commission, FTC Proposes New Protections to Combat AI Impersonation of Individuals (Feb. 15, 2024), https://www.ftc.gov/news-events/news/press-releases/2024/02/ftc-proposes-new-protections-combat-ai-impersonation-individuals [https://perma.cc/6FME-F4CY].

8 *See* Federal Trade Commission, Comment Letter on U.S. Copyright Office Notice of Inquiry on Artificial Intelligence and Copyright 5–6, (Oct. 30, 2023), https://www.regulations.gov/comment/COLC-2023-0006-8630 [https://perma.cc/8D4N-GZC5].

9 I address the *prima facie* infringement claim in a contribution to *Nimmer on Copyright*. *See* 5 MELVILLE B. NIMMER & DAVID NIMMER, NIMMER ON COPYRIGHT § 20.05[C][1] (2024). Some scholars reject this framing, arguing that we need not reach fair use because copyright should not recognize training copies as actionable. *See* Oren Bracha, *The Work of Copyright in the Age of Machine Production*, 38 HARV. J.L. & TECH. 171, 181 (2024) ("[T]raining copies involve no reproduction of copyrightable subject matter and therefore cannot infringe."); Carys J. Craig, *The AI-Copyright Trap*, 100 CHI.-KENT L. REV. (forthcoming 2025) (manuscript at 22–23) (July 15, 2024 draft on file with author) (dismissing training copies as "immaterial both literally and figuratively"); *see also* Michael D. Murray, *Generative AI Art: Copyright Infringement and Fair Use*, 26 SMU SCI. & TECH. L. REV. 259, 285–86 (2023) (arguing that training does not factually involve copying). The typical argument in favor of fair use is that training is transformative, *see, e.g.*, Pamela Samuelson, *Fair Use Defenses in Disruptive Technology Cases*, 71 UCLA L. REV. 1484, 1558 ("Insofar as generative AI systems' uses of in-copyright works are for very different purposes than the originals' . . . the AI training uses are likely to be considered transformative."), or non-expressive, *see, e.g.*, Matthew Sag, *Copyright Safety for Generative AI*, 61 HOUS. L. REV. 295, 308 (2023) (positing "there is no reason to think that courts would, or should, apply the principle of nonexpressive use differently to text data mining when it is used in machine learning"). Counterarguments take the opposite side, *see, e.g.*, Benjamin L.W. Sobel, *Artificial Intelligence's Fair Use Crisis*, 41 COLUM. J.L. & ARTS 45, 70 (2017) ("[C]omputerized

Allegations of intellectual theft and matters regarding the creation of new works are squarely within copyright's wheelhouse, and copyright has been the *de facto* regulator of new media technologies since almost its inception.[10] Copyright's fair-use defense has been especially salient with respect to technologies of mass copying, from the photocopier through more recent mass-digitization projects.[11] Over twenty copyright suits against AI companies now wind through the courts[12] while the Copyright Office sifts through over ten thousand public submissions to its Notice of Inquiry regarding AI copyright.[13] Commentators have declared that "AI's future could hinge on one thorny legal question"—that of fair use.[14]

This framing is wrong but nonetheless productive. It is wrong because fair use is not generative AI's live-or-die question. The framing is productive, however, because fair-use decisions pose deep doctrinal and normative challenges for copyright[15] and because the implications for control over these systems will undoubtedly matter for our collective cultural

consumption of authorial expression might also constitute infringement if that consumption implicates the expressive value in those works."), or question the premises of non-expressive use, *see* David W. Opderbeck, *Copyright in AI Training Data: A Human-Centered Approach*, 76 OKLA. L. REV. 951, 976 (2024) (questioning "the theoretical and practical basis for this supposed doctrine"); Robert Brauneis, *Copyright and the Training of Human Authors and Generative Machines*, 47 COLUM. J.L. & ARTS (forthcoming 2025) (July 31, 2024 draft on file with author).

10 *See generally* Jane C. Ginsburg, *Copyright and Control Over New Technologies of Dissemination*, 101 COLUM. L. REV. 1613 (2001); Blake E. Reid, *What Copyright Can't Do*, PEPP. L. REV. (forthcoming 2025) (Sept. 8, 2024 draft on file with author).

11 *See* 4 NIMMER & NIMMER, *supra* note 9, § 13F.13 (detailing fair-use jurisprudence for analog technologies); *id.* § 13F.14 (detailing fair-use jurisprudence for digital technologies); Matthew Sag, *Copyright and Copy-Reliant Technology*, 103 NW. U. L. REV. 1607 (2009) (tracing fair use through the Internet age); Fred von Lohmann, *Fair Use as Innovation Policy*, 23 BERKELEY TECH. L.J. 829 (2008) (detailing copyright's interplay with innovation).

12 *See Status Report on All 24 Copyright Lawsuits v. AI Companies*, CHAT GPT IS EATING THE WORLD (June 6, 2024), https://chatgptiseatingtheworld.com/2024/06/06/status-report-on-all-24-copyright-lawsuits-v-ai-companies-jun-6-2024-j-l-v-alphabet-dismissed-nvidia-hires-neal-katyal [https://perma.cc/9LD9-QVXW].

13 Letter from Shira Perlmutter, Reg. of Copyrights & Dir., U.S. Copyright Office, to Senator Chris Coons, Chair, and Senator Thom Tillis, Ranking Member, Subcomm. on Intell. Prop. (Feb. 23, 2024), https://www.copyright.gov/laws/hearings/USCO-Letter-on-AI-and-Copyright-Initiative-Update.pdf [https://perma.cc/GC4Y-FDDK].

14 Will Oremus & Elahe Izadi, *AI's Future Could Hinge on One Thorny Legal Question*, WASH. POST (Jan. 4, 2024), https://www.washingtonpost.com/technology/2024/01/04/nyt-ai-copyright-lawsuit-fair-use [https://perma.cc/Z6LV-QJT8].

15 On the relation between the application and normative uncertainties posed by new technologies, see Rebecca Crootof & BJ Ard, *Structuring Techlaw*, 34 HARV. J.L. & TECH. 347, 356–57 (2021).

engagement[16] and the state of competition in the burgeoning AI industry.[17] Yet generative AI will be deployed regardless, illustrating deeper structural limitations of the copyright regime.[18] Several developers have already built functional image-generation systems without recourse to unauthorized copying.[19] Training for these systems stands outside copyright enforcement, pointing to the need for other mechanisms to vindicate several of the ends we normally entrust to copyright.[20]

The analysis proceeds in four parts. Part I provides a technical introduction to generative AI, grounded in the operation of image systems like OpenAI's DALL-E 2.[21] Because we are investigating fair use as to copying during training, the discussion begins with a start-to-finish primer for non-technical audiences, moving from the use of existing images to train AI models to the use of trained models to create new images.[22] Central to this explanation is OpenAI's creation of a "latent space" that encodes all images DALL-E 2 is capable of making into a shared spatial representation. Its training process culminated in the creation of one model that established the coordinates for this space, another that selects coordinates to match the user's prompt, and yet another to decode those coordinates into images with the desired features.

Doctrinal and practical reasons compel us to go beyond training, however, to understand the "supply chain" linking AI models to completed systems[23] and to examine the systems'

---

16 *See infra* subpart IV.A.

17 *See infra* subpart IV.C.

18 *See* Reid, *supra* note 10, at 51–55 (detailing this argument).

19 *See infra* subpart IV.B.

20 *Cf.* Craig, *supra* note 9 (arguing that invoking copyright to protect individual artists may instead aid powerful interests).

21 I completed a first draft of this Article when DALL-E 2 was cutting edge and the first copyright lawsuits against OpenAI were mere months old. "Copyright-safe" systems like Adobe Firefly had not yet launched, nor had today's systems that can transform text to video. At the time of this publication, OpenAI has discontinued DALL-E 2 and redirected users to the image-generation systems integrated into ChatGPT. The deep dive into DALL-E 2 nonetheless holds value. Hardly any primers describe image-generation systems in detail for a non-technical audience and understanding an early system like DALL-E 2 provides the scaffolding for making sense of the different capabilities and design decisions of more complex systems. Understanding the range of possibilities for how an AI model might be trained or an AI system might be deployed helps us better grapple with the doctrinal and theoretical puzzles these systems pose.

22 *See infra* subparts I.B–I.C.

23 *See infra* subpart I.A. Lee, Cooper, and Grimmelmann provide an extraordinary introduction and explanation of the AI supply-chain framework in Katherine Lee, A. Feder Cooper & James Grimmelmann, *Talkin' 'Bout AI Generation:*

outputs.[24] Fair use conventionally requires assessing the purposes and effects of copying.[25] This was straightforward for most prior copying technologies: purpose was evident from the tight relationship between preliminary copying and the ultimate system, and the effects of exploiting any given work were easy to trace. Neither point holds true for generative AI. First, an AI model's purposes may be indeterminate at the time of training.[26] Training is merely one step on the supply chain through which developers create, configure, and combine models to create user-facing AI systems. The model creator need not be the party who later configures it into a system, and the model's capabilities and uses may be indeterminate until that final stage. Second, AI systems may or may not store or reproduce recognizable pieces of existing works.[27] Our inability to directly scrutinize AI models or a system's latent space leaves us to infer a system's contents and capabilities from its outputs.

Part II works through the puzzles that arise under fair use's conventional tests. Transformative use hinges on the purpose of the use[28]—but at what juncture in a multi-step, multi-party supply chain do we assess the purpose of training?[29] Other approaches to fair use hinge on whether the use substitutes for the original[30]—how do we weigh the competitive harm of an AI model that displaces artists by replicating uncopyrightable elements of prior works? Then there is the theory that non-expressive uses are fair.[31] To qualify as non-expressive use, is

---

Copyright and the Generative-AI Supply Chain, 68 J. COPYRIGHT SOC'Y U.S.A. (forthcoming 2025) (Mar. 1, 2024 draft on file with author).

24 *See infra* subpart I.D.

25 *See infra* subpart II.A–II.C.

26 *See infra* subpart I.A.

27 *See infra* subpart I.D.

28 *See infra* subpart II.A.

29 *Compare infra* section II.B.1 (detailing an "entanglement approach" that links the purpose of preliminary copying with that of the final product) *with infra* section II.B.2 (detailing a "disaggregation approach" that examines each act of copying in isolation).

30 *See infra* subpart II.C (detailing the *Warhol* Court's "substitutability test").

31 *See infra* subpart II.D. The critical question in the original formulation of non-expressive fair use is whether the use communicates an author's original expression to the public. *See* Matthew Sag, *The New Legal Landscape for Text Mining and Machine Learning*, 66 J. COPYRIGHT SOC'Y U.S.A. 291, 314–19 (2019) [hereinafter Sag, *New Legal Landscape*]; *see also* Sag, *supra* note 11. Theories like fair learning, *see* Mark A. Lemley & Bryan Casey, *Fair Learning*, 99 TEX. L. REV. 743, 772 (2021), the freedom to extract unowned elements, *see* Molly Shaffer Van Houweling, *The Freedom to Extract in Copyright Law*, 103 N.C. L. REV.

it sufficient that the final system avoids replicating expressive elements from the training works, or must the training process itself be indifferent to the works' expressive content?[32]

These questions lingered prior to the advent of AI, but they previously stood unanswered because the distinctions seldom mattered. Most cases could be decided the same way under any of the competing theories. AI has brought fresh urgency because different choices may lead to divergent outcomes. Unpacking the tensions reveals the limits of the transformative-use paradigm[33] and the need for more rigorous examination of which kinds of market harm count.[34]

Part III makes headway on these debates by pulling from copyright's latent space. The move is figurative—unlike an AI system, copyright lacks a literal multi-dimensional space in which cases are encoded and from which we can interpolate legal principles. But copyright jurisprudence does possess a set of incompletely articulated commitments that come into view only when we ask the right questions. Here, a focus on permissible substitutions allows articulation of a new principle: the freedom to exploit a work's non-authorial value notwithstanding potential market harm.[35] This framing sidesteps the transformative-purpose question, though it suggests that uses that are simultaneously transformative *and* not exploitative of authorial value are especially likely to be fair. It bolsters the position that competition on the basis of copying non-expressive or otherwise unprotected elements should not count: copyright does not regard the value of facts and other unprotected elements as originating with the work's author.[36] And it augments theories of non-expressive use by arguing that copying some forms of expression—including tropes and scenes-a-faire elements, and sometimes newsworthy materials—may be fair because the value of these elements derives from societal expectations and third-party interests.[37] But the principle should not be taken as a free pass. Caselaw also shows that drawing the line between what is authorial and what is not is

(forthcoming 2025) (Oct. 28, 2024 draft on file with author), and the value of spillovers, *see* Bracha, *supra* note 9, at 179–81, provide complementary perspectives.

32 *See* Brauneis, *supra* note 9 (developing these distinctions as part of a larger taxonomy); *see also infra* section II.D.2 (unpacking that taxonomy).

33 *See infra* section II.B.3.

34 *See infra* subparts II.C–II.D.

35 *See infra* subpart III.A.

36 *See infra* subpart III.B.

37 *See infra* subpart III.C.

not an exercise in conceptual purity, but instead a pragmatic exercise grounded in the reconciliation of copyright policy and market realities in specific contexts.[38] Copyright's policy response to generative AI remains to be hashed out.

All that said, working through the fair-use question illustrates why fair use cannot stand as the mechanism for mitigating AI's anticipated harms. Part IV explains that the problem is partly one of reconciling contested values. Some would argue fair use should be denied because public policy demands it.[39] Yet that conclusion is far from settled. The argument that copyright policy requires protecting artists from AI must contend with the counterargument that AI advances other copyright interests like semiotic democracy.[40] Counterarguments about the importance of democratizing cultural production must likewise grapple with open questions about which groups benefit from upholding or denying fair use.[41] And then there are conflicting positions on whether it is appropriate to conscript copyright to mitigate policy concerns outside the realm of copyright, like the harms of deepfakes and tech monopolization.[42]

Part IV concludes by arguing fair use is the wrong regulatory mechanism no matter how we define copyright's policy commitments. Copyright's fundamental structural limitations render it incapable of vindicating the stated concerns.[43] Even assuming consensus that generative AI should be stopped, copyright comes into play only in the event of unauthorized copying. Existing image systems demonstrate that it is entirely possible to create outputs that displace established artists without copying those artists' work.[44] The core problem is that generative

38 *See infra* subpart III.D.

39 *See infra* subpart IV.A.

40 For insight into that debate, see generally Katrina Geddes, Generative AI's Public Benefit (Feb. 1, 2025) (unpublished manuscript), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4865510 [https://perma.cc/AR7U-YWFE] (detailing displacement arguments alongside AI's potential to advance semiotic democracy).

41 *See* Craig, *supra* note 9, at 24–26 (explaining how *denying* fair use may privilege entrenched interests); Benjamin L.W. Sobel, Copyright Accelerationism 31–34 (Dec. 8, 2023) (unpublished manuscript), https://ssrn.com/abstract=4658701 [https://perma.cc/7WK3-TT56] (explaining how *extending* fair use may privilege entrenched interests).

42 *See* Matthew Sag, *Fairness and Fair Use in Generative AI*, 92 FORDHAM L. REV. 1887, 1899 (2024) (arguing fair use should not turn on "broader public interest arguments").

43 *See infra* subpart IV.B; *see also* Reid, *supra* note 10.

44 *See* Rashi Shrivastava, *Adobe Brings Its Generative AI Tool Firefly to Businesses*, FORBES (June 8, 2023), https://www.forbes.com/sites/

systems can produce comparable works cheaply and quickly.[45] Denying fair use would not protect these artists from competition.[46] It could nonetheless exacerbate the problem of tech monopolization by advantaging platforms like Meta that are positioned to exploit the troves of data they have lawfully extracted from their users.[47] Meeting the challenges that generative AI poses for the future of art and the rest of society requires a coordinated response through which copyright is but one tool alongside others for promoting the arts, blunting the misuse of AI, and confronting the technology's capacity to consolidate power.

I
State of the Art

We can better articulate the challenges generative AI poses for copyright if we begin with working knowledge of the technology. This Part uses OpenAI's DALL-E 2 as a case study to probe a generative image system's training and operation.[48]

---

rashishrivastava/2023/06/08/adobe-brings-its-generative-ai-tool-firefly-to-businesses [https://perma.cc/7V9B-7VKJ]; Press Release, Getty Images, Getty Images Launches Commercially Safe Generative AI Offering (Sept. 25, 2023), https://newsroom.gettyimages.com/en/getty-images/getty-images-launches-commercially-safe-generative-ai-offering [https://perma.cc/XE6S-RPS7] [hereinafter Getty Press Release]; Edwards, *supra* note 4. Alternative strategies like the use of synthetic training data may also multiply the availability of quality non-infringeable training works. *See generally* Peter Lee, *Synthetic Data and the Future of AI*, 110 Cornell L. Rev. 1 (2025).

45 *Cf.* Bracha, *supra* note 9, at 223 (arguing the underlying complaint falls outside copyright's concerns with copying of discrete works and instead concerns the uses and impacts of aggregate metainformation).

46 Conversely, *granting* fair use would do only so much even if the consensus were that widespread availability of generative AI advanced copyright policy. It could remove the specter of copyright liability for training, but it would not compel developers to create pro-social AI systems, *cf.* Reid, *supra* note 10, at 55–57, nor would it remove entrenched interests' incentives to over-filter the resulting systems, *see* Geddes, *supra* note 40, at 69–71.

47 *See infra* subpart IV.C.

48 Focusing on images puts us at the vanguard of legal action: Image systems preceded text systems in copyright litigation. Individual artists filed suit against the owners of the image systems Midjourney and Stable Diffusion in January 2023, *see* Complaint, Andersen v. Stability AI, Ltd., No. 3:23-cv-00201, 700 F. Supp. 3d 853 (N.D. Cal. filed Jan. 13, 2023), and Getty Images followed with its own suit against Stable Diffusion in February 2023, *see* Complaint, Getty Images (US), Inc. v. Stability AI, Ltd., No. 1:23-cv-00135 (D. Del. filed Feb. 3, 2023). Authors did not initiate infringement involving ChatGPT and other language systems until June and July. *See* Complaint, Tremblay v. OpenAI, Inc., No. 3:23-cv-03223 (N.D. Cal. filed June 28, 2023); Complaint, Silverman v. OpenAI, Inc., No. 3:23-cv-03416 (N.D. Cal. filed July 7, 2023). Image systems also stood at the forefront of proceedings regarding the copyrightability of AI outputs, *see* Thaler v. Perlmutter, 687 F. Supp. 3d 140 (D.D.C. Aug. 18, 2023) (appeal filed Oct. 18, 2023),

Each AI system utilizes one or more AI models—DALL-E 2 uses three.[49] The training process for each model aims to map information about the images used for training to a spatial representation of information known as a latent space.[50] Training assigns information to distinct points in the latent space called "embeddings."[51] The resulting latent space is one in which every possible embedding corresponds to a set of image features.[52] Although this approach to organizing information may seem convoluted, the latent space allows us to explain a system like DALL-E 2 in simple terms: the system uses one model to encode the user's text prompt into the system's language, a second model to select an embedding based on the encoded prompt, and a third model to decode that embedding into a new image.

Challenges arise because the copying that occurs during a model's training is distant from the final system's operation. Some tests for fair use focus on the purpose of copying.[53] However, an AI model's purpose may be indeterminate or contingent at the time it is trained.[54] The initial training is but one step in a longer "generative-AI supply chain" that culminates in the completed AI system, and the model's intended or realized uses may depend on subsequent rounds of training, the model's

---

and the development of licensed alternatives, *see supra* note 44 and accompanying text.

49 *See infra* subpart I.C. OpenAI details the system in the DALL-E 2 Paper, *supra* note 3. DALL-E 2 combines OpenAI's prior GLIDE system, Alex Nichol et al., *GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models*, ARXIV (Mar. 8, 2022), https://arxiv.org/pdf/2112.10741 [https://perma.cc/YXZ4-7BWL] [hereinafter GLIDE Paper], with its CLIP model, Alec Radford et al., *Learning Transferable Visual Models From Natural Language Supervision*, ARXIV (Feb. 26, 2021), https://arxiv.org/pdf/2103.00020 [https://perma.cc/W6WS-GVX5] [hereinafter CLIP Paper]. For its part, CLIP was originally developed as an accessory to DALL-E 1. *Id.* at 6; *see* Aditya Ramesh et al., *Zero-Shot Text-to-Image Generation*, ARXIV 6 (Feb. 26, 2021), https://arxiv.org/pdf/2102.12092v2 [https://perma.cc/5J97-666F] [hereinafter DALL-E 1 Paper].

50 *See infra* subpart I.C.

51 *See* Joel Barnard, *What Is Embedding?*, IBM (Dec. 22, 2023), https://www.ibm.com/topics/embedding [https://perma.cc/29VT-VF6F].

52 *See* Ian Stenbit, François Chollet & Luke Wood, *A Walk Through Latent Space with Stable Diffusion*, KERAS (Sept. 28, 2022), https://keras.io/examples/generative/random_walks_with_stable_diffusion [https://perma.cc/PX2Y-4266] (producing different images by "walking" through latent space). Although each embedding produces an image, the results may prove nonsensical or horrific if the embedding sits outside the portion of the latent space mapped through training. *See* Devin Coldewey, *A Terrifying AI-Generated Woman is Lurking in the Abyss of Latent Space*, TECHCRUNCH (Sept. 13, 2022), https://techcrunch.com/2022/09/13/loab-ai-generated-horror [https://perma.cc/7UUH-UFBX].

53 *See infra* subparts II.A–B.

54 *See infra* subpart I.A.

combination with other, separately trained models, and design decisions with respect to the configuration of the completed system.[55] Adding to the distance between training and ultimate use, these later steps may be undertaken by parties other than the one who trained the model.[56]

Other tests for fair use depend on the system's capabilities and the resulting impacts.[57] The inquiry begins with a deceptively simple question: what sorts of images can one extract from the latent space? As to impact, do the images reproduce or compete with images copied during training? The problem is that we cannot inspect a latent space or the underlying AI models directly; our best inferences depend on examining the system's outputs.[58] Here we can see that—depending on decisions during training and in overall system design—image systems have the capability to memorize and regurgitate images verbatim and also the potential to generalize and create things distinct from items in the training set.[59] Complications follow where generalization leads to things that resemble or compete with training works without legally or factually copying them.[60]

A. Systems Thinking

The relation between systems and models is crucial to analyzing generative AI. Systems are complete products like DALL-E 2 and ChatGPT.[61] Because users typically interact with systems rather than models, most discussions regarding AI's impacts are best understand as conversations about the capabilities and consequences of systems.[62] Models are subcomponents that are crucial to a system's function: each AI system works by configuring one or more AI models to work together or with specific software.[63] Models are also central in copyright analysis—copying during training is a key concern, and training is coextensive with

---

55 *See id.*; *see also infra* section I.D.1.

56 *See infra* subpart I.A.

57 *See infra* subpart II.C (detailing tests centered on competitive consequences); *infra* subpart II.D (detailing tests centered on the reproduction of protected expression).

58 *See infra* subpart I.D.

59 *See id.*

60 *See id.*

61 Artificial Intelligence and Copyright, 88 Fed. Reg. 59942, 59948 (Aug. 30, 2023).

62 *See* Lee, Cooper & Grimmelmann, *supra* note 23, at 16–17. "A model by itself is an inert artifact." *Id.* at 5.

63 88 Fed. Reg. at 59948.

model creation.[64] Each AI model is trained to perform a specific task or set of tasks determined by the creator's selection of inputs, training algorithms, and model architectures.[65]

Lee, Cooper, and Grimmelmann detail a "generative-AI supply chain" tracing the steps for creating a model and configuring it into a system.[66] Preliminary activity goes toward assembling works for training,[67] which may be no small feat given that current-generation systems may train on millions or billions of works.[68] The first round of training—sometimes called "pre-training"—deploys machine learning on a large corpus of works to create what is known as a "pre-trained model"[69] or "foundation model."[70] Subsequent training rounds—often called "fine-tuning"—utilize a more targeted body of works to refine the model.[71] Only after training is complete do system designers configure the trained models into a system.[72] The supply chain metaphor speaks to the fact that each link on the chain is a separate round of activity that may be conducted by a separate party.[73] Indeed, some parties seek to promote AI development by undertaking the costly pre-training process and distributing their models for others to fine-tune.[74]

The pretrained model's capabilities may be general yet limited prior to fine-tuning and incorporation into a system. As case in point, the core function of the GPT model behind ChatGPT is simply to predict the next word in a sequence.[75] Systems like ChatGPT can perform more elaborate feats owing to fine-tuning,

---

64 *See* Lee, Cooper & Grimmelmann, *supra* note 23, at 12–13.

65 *See id.* at 11–12.

66 *Id.* at 4–5; *see also* Paul Ohm, *Focusing on Fine-Tuning: Understanding the Four Pathways for Shaping Generative AI*, 25 COLUM. SCI. & TECH. L. REV. 214, 219–31 (2024) (detailing four steps in the process and possible interventions at each stage).

67 Lee, Cooper & Grimmelmann, *supra* note 23, at 5.

68 *See supra* notes 3–4 and accompanying text.

69 Lee, Cooper & Grimmelmann, *supra* note 23, at 4, 41–42.

70 *See generally* Peter Henderson et al., *Foundation Models and Fair Use*, 24 J. MACH. LEARNING RSCH., Sept. 2023, at 2, 5.

71 Lee, Cooper & Grimmelmann, *supra* note 23, at 42–43.

72 *Id.* at 45–49.

73 *Id.* at 147 ("Every single one of these steps could be under the control of a different person."); *see generally* Marcela S. Melara & Mic Bowman, *What is Software Supply Chain Security?*, ARXIV (Sept. 8, 2022), https://arxiv.org/pdf/2209.04006 [https://perma.cc/DGZ9-VMXE] (detailing involvement of multiple parties and reliance on third parties as key challenges for software supply chains).

74 *See* Ohm, *supra* note 66, at 228.

75 *See* Artificial Intelligence and Copyright, 88 Fed. Reg. 59942, 59948 (Aug. 30, 2023).

the incorporation of further models, and the addition of other software.[76] Moreover, pre-trained models may be capable of either generative or non-generative uses.[77] Subject to fine-tuning, the same image-to-text models that map a latent space for generative AI can enable screen-reading systems that translate images into textual descriptions for people with visual impairments.[78]

Although models are central to the function of AI systems, the choices that determine a system's performance go beyond model selection and refinement.[79] Indeed, systems using identical models may perform quite differently because of other design decisions. Consider two separate systems built using an AI model with the technical capacity to produce any image on command. The creator of the first system might program it to accept all prompts indiscriminately, unleashing the model's full capabilities. The creator of the other system might program it to reject prompts naming public figures or copyrighted characters and to reject prompts describing violence or sexual activity. The latter system performs quite differently not because of any difference in the models—they are identical—but because of other system-configuration choices.

B. Machine Learning Basics

Understanding how copyrighted works are used in model creation requires a brief introduction to the machine learning processes through which models are trained.[80] Machine learning is a catch-all term for techniques through which developers direct a computer to process information to achieve the

---

76 *See* Ohm, *supra* note 66, at 223–24.

77 Henderson et al., *supra* note 70, at 4.

78 CLIP Paper, *supra* note 49, at 20 ("Many of CLIP's capabilities are omni-use in nature (e.g. OCR can be used to make scanned documents searchable, to power screen reading technologies, or to read license plates)."); *see also* Khari Johnson, *AI Could Change How Blind People See the World*, WIRED (July 5, 2023), https://www.wired.com/story/ai-gpt4-could-change-how-blind-people-see-the-world [https://perma.cc/R6PY-H8C2] (explaining how generative AI powers emerging assistive systems); Tianqi Wei, Zhi Chen & Xin Yu, *Snap and Diagnose: An Advanced Multimodal Retrieval System for Identifying Plant Diseases in the Wild*, ARXIV 1 (Aug. 27, 2024), https://www.arxiv.org/pdf/2408.14723 [https://perma.cc/RQ6V-7BXY] (proposing a system that diagnoses plant diseases by incorporating a model conventionally used for image systems).

79 Lee, Cooper & Grimmelmann, *supra* note 23, at 147; *see also infra* subpart I.D.1.

80 The term "machine learning" is not without controversy. Although these are the terms of art used in computer science, some may object to discussions of "learning," and processes of "training" or "inference," for attributing human-like cognition to these systems or insinuating that these systems should be subject to the same rules as humans. I use the term not to endorse anthropomorphic thinking, but instead to enter the dialogue using standard terminology.

capacity to complete a particular task; this process is called "training."[81] These techniques are fundamental to generative AI.[82] AI developers often engage in successive training rounds, first to extract information from a body of works, often called a training set, and later to devise processes to create new works using that information.[83]

Training reflects a "guess-and-adjust" strategy whereby the model creator repeatedly processes the training data to set and update a model's parameters.[84] To better understand this strategy, consider a concrete application in the text modality. The process for training a language model might involve constructing sentences like "Sally and Jesse brought sunscreen to the __."[85] In guess-and-check fashion, a computer would test different predictions for the blank, adjusting model parameters to increase the likelihood that the model would select a word that often appears near "sunscreen" in the training set.[86] Assuming a successful training process, the resulting language model would reflect that "beach" frequently appears in this context but "dentist" does not.

The process of creating AI models for image generation begins with analogous training to map relationships between

---

81 *See* 88 Fed. Reg. at 59949.

82 *Id.*

83 *See infra* subpart I.C.

84 Lee, Cooper & Grimmelmann, *supra* note 23, at 14–15. This sets machine learning applications in contrast with traditional computer programming and raises distinct legal issues. A programmer who wished to manually create an algorithm for sorting job applicants might come up with an explicit rubric for scoring applicants on the basis of, say, degrees, years of experience, and answers to a questionnaire. By contrast, an engineer using machine learning could feed a system top employees' resumes to identify success markers.

Legal scholars have flagged many problems with machine learning, particularly its propensity to replicate biases in the data. For example, overrepresentation of privileged groups in existing resumes led one system to screen applicants for markers including high school lacrosse and the name "Jared." Lori Andrews & Hannah Bucher, *Automating Discrimination: AI Hiring Practices and Gender Inequality*, 44 CARDOZO L. REV. 145, 154–55 (2022). Concerns with misuse of race and proxies for race have plagued systems trained for tasks like criminal sentencing. *See* Ngozi Okidegbe, *Discredited Data*, 107 CORNELL L. REV. 2007 (2022). These concerns are compounded by transparency and explainability problems: machine learning systems do not typically store inferences in a manner intelligible to humans. *See* Hannah Bloch-Wehba, *Access to Algorithms*, 88 FORDHAM L. REV. 1265, 1312–14 (2020). It is thus no surprise that image systems have also replicated biases, *see* Ohm, *supra* note 66, at 239–40, and that the unintelligibility of AI models presents complications for copyright, *see infra* subpart I.D.

85 *See* Meta Platforms, Inc., Comment Letter on U.S. Copyright Office Notice of Inquiry on Artificial Intelligence and Copyright 3–6 (Oct. 30, 2023), https://www.regulations.gov/comment/COLC-2023-0006-9027 [https://perma.cc/R9KR-ZA6G].

86 *See id.*

images and their captions.[87] Instead of identifying which words belong together, the computer engages in guess-and-adjust processes to determine which sorts of textual descriptions tend to accompany which sorts of pictures.[88] Through that process, the model becomes capable of determining that night-sky pictures correspond to captions like "millions of stars" and "Milky Way" rather than plausible competitors like "white dots on dark background." The model can do this because it is not trying to identify objectively accurate captions, but instead to match pictures with the language humans use to describe them.[89] Herein lies the need for large numbers of works: the AI models for the leading art systems ingested hundreds of millions of works to map the correspondence between words and images.[90]

C. From Training to Latent Space

1. *Modeling a Latent Space*

Our clearest window into how an AI model organizes information is its latent space: the spatial representation of correspondences a model draws between items.[91] The model itself is inscrutable because machine learning processes typically do not process or store information in formats comprehensible to humans.[92] Instead, models represent each training item as a set of coordinates called an embedding.[93] A map emerges when one documents the coordinates that a trained model assigns to different items. Recall the hypothetical language model trained to determine which things belong together (sunscreen and beaches) and which do not (sunscreen and dentists). We would find that it placed similar items like "sunscreen" and "beach" near one another on the virtual map. It would place dissimilar

---

87 *See infra* subpart I.C.1. In a later phase of training, model creators engage in a separate guess-and-adjust process to train a diffusion model to identify the mathematical functions to remove static from a "noised" picture to restore the original picture. *See infra* subpart I.C.3.

88 *See generally* CLIP Paper, *supra* note 49, at 4.

89 *See id.* at 3 (documenting advantages of using natural language).

90 OpenAI trained a model on 650 million image-text pairs to identify these correspondences, *see infra* section I.C.1, and utilized only 250 million pairs to train a model to draw new images, *see infra* section I.C.3.

91 *See* Yang Liu, Eunice Jun, Qisheng Li & Jeffrey Heer, *Latent Space Cartography: Visual Analysis of Vector Space Embeddings*, 38 COMP. GRAPHICS F. 67, 67–68 (2019).

92 Alicia Solow-Niederman, *Administering Artificial Intelligence*, 93 S. CAL. L. REV. 633, 657–58 (2020).

93 *See* Lee, Cooper & Grimmelmann, *supra* note 23, at 8–9.

items like “dentist” away from vacation-themed words, but close to “tooth” and “drill.” We could also prompt the model to assign coordinates to words not in the training set, including made-up words or misspellings: a model that internalized the relevant linguistic patterns might group the made-up word “devolf” with “wild beast” and “hellhound.”[94] Plotting these coordinates relative to one another yields a map of the model's latent space.

The machine learning process for an image system constructs a similar map with images. We could imagine, for example, a simple latent space organizing dog, bird, and airplane images into related clusters:

Figure 1: Visualizing a Simple Latent Space

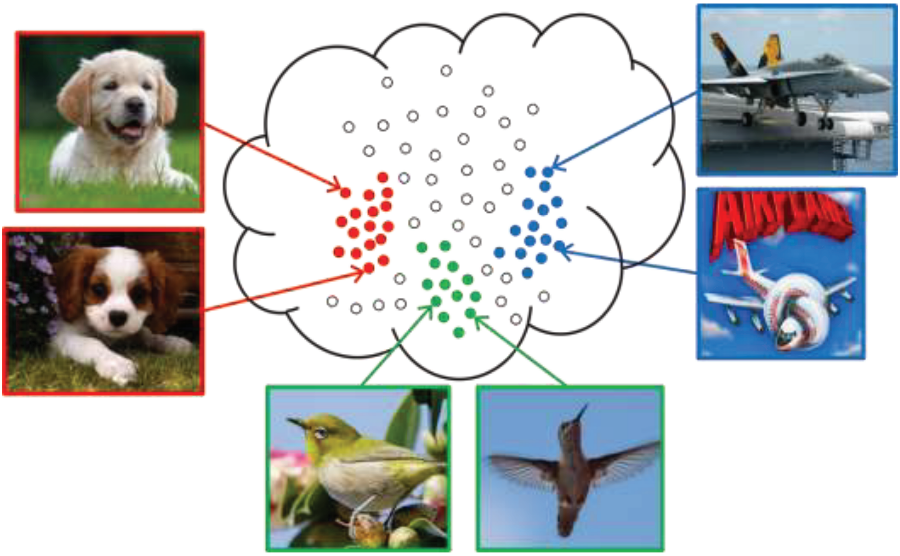

Figure 1: A latent space groups image embeddings by similarity. Placement of the bird cluster (green) between dogs (red) and airplanes (blue) is intentional, and reflects that birds are similar to dogs in some ways (*e.g.*, they have eyes and sometimes appear with vegetation) and to airplanes in others (*e.g.*, they have wings and sometimes appear in the sky). The latent space does not literally save or organize images. Instead, it represents each item as a unique embedding, a representation of the image as a list of numbers. Each circle here represents the embedding for one image.[95]

---

94 Mohammed Terry-Jack, *NLP: Everything About Embeddings*, MEDIUM (Apr. 21, 2019), https://medium.com/@b.terryjack/nlp-everything-about-word-embeddings-9ea21f51ccfe [https://perma.cc/UW3V-NQQ2].

95 This Article is available in full color via the author's SSRN page at https://ssrn.com/abstract=4630085 and via the *Cornell Law Review* website.

The dogs–birds–airplanes model is of course a vast simplification. A two- or three-dimensional cloud provides an intuitive mental model for organizing a handful of categories based on similarity. But a three-dimensional space cannot faithfully capture all the ways in which images and their captions might relate to one another. Picasso's cubist painting of a bull in "Guernica" relates only loosely to realistic depictions of bulls but resembles his black-and-white Cubist portraits of other subject matter. A close-up photo depicting actor Chris Hemsworth as Thor relates conceptually to comic-book drawings of Thor and also resembles photos of Chris Hemsworth on the red carpet. We could in theory account for extra layers of similarity and difference by adding additional axes to our map—which is what these models do.

OpenAI created the latent space for DALL-E 2 by training the AI model "CLIP," short for "Contrastive Language–Image Pretraining," on 650 million image-caption pairs.[96] The training process did not seek merely to assign embeddings directly to each of the 650 million items (*e.g.*, to place birds, dogs, and airplanes in separate piles). Instead, OpenAI conducted intensive guess-and-check work to devise a mathematical function that would assign each item to an embedding near the embeddings for similar items.[97] CLIP ultimately obtained the capacity to convert each image into a unique image embedding that specifies a latent-space position along 319 dimensions.[98] In parallel, it also converted each caption to a text embedding in a

96 DALL-E 2 paper, *supra* note 3, at 23. The text captions played two important roles. First, they provided guidance on which images were similar. As pixels on a screen, pictures captioned "dog in grass" and a "bird in reeds" may be more similar than the pictures "bird in reeds" and "bird in sky;" the captions allow the machine learning process to home in on the similarities that make the bird-captioned pictures similar despite visual differences. *See generally* CLIP Paper, *supra* note 49. Second, CLIP devised a latent space for the captions themselves. Comparing the relative position of an image embedding, in the image latent space, and a caption embedding, in the text latent space, allows for determining whether the caption matches the image. *See id.*

97 As noted above with the "devolf," successful training will result in an AI model that can place items from *outside* the training set at reasonable locations between coordinates that correspond to training items.

98 *See* DALL-E 2 paper, *supra* note 3, at 3–4. Those who are familiar with converting a file into a "hash" for purposes of de-duplication in e-discovery, or for purposes of working with a blockchain, may recognize the embedding as playing a similar identifier function. A hash, however, is arbitrary relative to the document it represents; changing a single word in a document can generate an entirely different hash value. Embeddings are different because AI models assign similar embeddings to similar items.

parallel text latent space.[99] The goal, again, was not to create a sorting algorithm good only for the images and captions in the training set. The goal was to devise a function that could assign embeddings to images the system had not previously seen.[100] This is what is remarkable about CLIP—faced with a new, uncaptioned photo of something like an orca, it will assign it an embedding within the same region as other killer-whale pictures.

Although CLIP is fundamental to DALL-E 2, it debuted as an auxiliary component to the original DALL-E ("DALL-E 1");[101] DALL-E 1 generated several images using AI models unrelated to CLIP, and the system invoked CLIP separately to determine which image best matched the user's prompt.[102] The user received only the image CLIP selected. Owing to this design, CLIP cannot generate anything on its own despite having mapped a comprehensive latent space. It can only select matches among existing items. Give it the caption "Disney cartoon crab" along with 10,000 pictures, and it will calculate and compare the associated embeddings to provide a definitive answer as to which of the 10,000 candidates possesses the image embedding closest to the caption's text embedding. If *The Little Mermaid*'s Sebastian is present, the system will likely select him. If not, the system will pick the nearest runner-up. But give it the caption alone and it cannot fabricate a new image embedding, let alone a new image. For that, additional models enter play.

2. *Prompt Translation*

An image latent space does not consist solely of embeddings for training images. With enough data points, an AI model can map a latent space with embeddings corresponding to an effectively infinite number of points beyond those associated with training items.[103] In effect, it infers what belongs in the empty

---

99 *See* GLIDE Paper, *supra* note 49, at 5.

100 CLIP Paper, *supra* note 49, at 6.

101 *Id.* at 6; *see* DALL-E 1 Paper, *supra* note 49, at 6–7.

102 *See* DALL-E 1 Paper, *supra* note 49, at 6–7 (citing the CLIP Paper, *supra* note 49, for its ranking mechanism). CLIP did so by comparing the position of the prompt's text embedding in the text latent space against the positions of the images' embeddings in the image latent space. *See* Grigory Sapunov, *OpenAI and the Road to Text-Guided Image Generation: DALL-E, CLIP, GLIDE, DALL-E 2 (unCLIP)*, MEDIUM (May 1, 2022), https://moocaholic.medium.com/openai-and-the-road-to-text-guided-image-generation-dall-e-clip-glide-dall-e-2-unclip-c6e28f7194ea [https://perma.cc/YJP4-9JYA].

103 *See supra* note 52 and accompanying text.

spaces. One advantage of organizing information this way is that it allows traversal of the space via mathematical reasoning. The chestnut example is the ability to navigate using word embeddings so that one can begin with the embedding for "king," direct the system to move *away* from the embedding for "man" and *toward* the embedding of "woman," and thereby arrive at an embedding for "queen" or a close approximation.[104]

Identifying these in-between points helps to clear a common misconception. People often imagine that generative art systems mash together preexisting works by calling two or more training images into memory and morphing them together as though the system was a sophisticated Instagram filter—Michael Murray dubs this misconception the "Magic File Drawer" theory.[105] This was the operative theory advanced and rejected early in the artists' lawsuit *Andersen v. Stability AI*.[106] But the truth is stranger. When asked to create a hybrid of two images, a system chooses an embedding meant to correspond to the desired image features and sends it to a decoder model.[107] The embedding will likely fall between the embeddings for training images corresponding to related image features. Contrary to the expectations of the Magic File Drawer theory, however, the system does not call any training images into memory. Nor does it need to. The pertinent models had already mathematically determined the embedding's image features when they completed their training months or years prior to producing the new image.[108]

---

104 *See* Alec Radford, Luke Metz & Soumith Chintala, *Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks*, ARXIV 8 (Jan. 7, 2016), https://arxiv.org/pdf/1511.06434v2 [https://perma.cc/H25C-U3ET]; *see also* DALL-E 2 paper, *supra* note 3, at 7–8. Similar permutations have revealed biases absorbed from training data: applying the same operations to the embedding for "doctor" yields "nurse" in some models. Timothy B. Lee & Sean Trott, *Large Language Models, Explained with a Minimum of Math and Jargon*, UNDERSTANDING AI (July 27, 2023), https://www.understandingai.org/p/large-language-models-explained-with [https://perma.cc/R6DD-FAVW]. These moves are possible because each embedding, as a string of numbers, is a vector susceptible to arithmetic manipulation. *See id.*

105 *See* Murray, *supra* note 9, at 302.

106 *See* Andersen v. Stability AI Ltd., 700 F. Supp. 3d 853 (N.D. Cal. 2023).

107 Diffusion decoders are explained at section I.C.3 below.

108 The point receives greater attention in discussions of memorization, where the point is that memorization happens at the time of training independent of whatever happens at the time of image generation. *See* A. Feder Cooper & James Grimmelmann, *The Files are in the Computer: On Copyright, Memorization, and Generative AI*, CHI.-KENT L. REV. (forthcoming 2025) (manuscript at 23) (July 22, 2024 draft on file with author).

To illustrate, consider that Baroque painter Johannes Vermeer seldom painted animals. He almost certainly never painted a rendition of "Girl with a Pearl Earring" featuring a marine mammal. But the information to produce an image corresponding to an oil painting featuring an otter in place of the girl exists, along the relevant dimensions, in a latent-space region between the embedding for Vermeer's original and the embeddings for various otter images. When DALL-E 2 renders the image, it does not recall pre-existing images and splice them. Instead, it identifies the closest matching point in the pre-constructed latent space and expresses it via its diffusion model:[109]

Figure 2: Intermediate Image Embeddings Exist Between Training-Image Embeddings

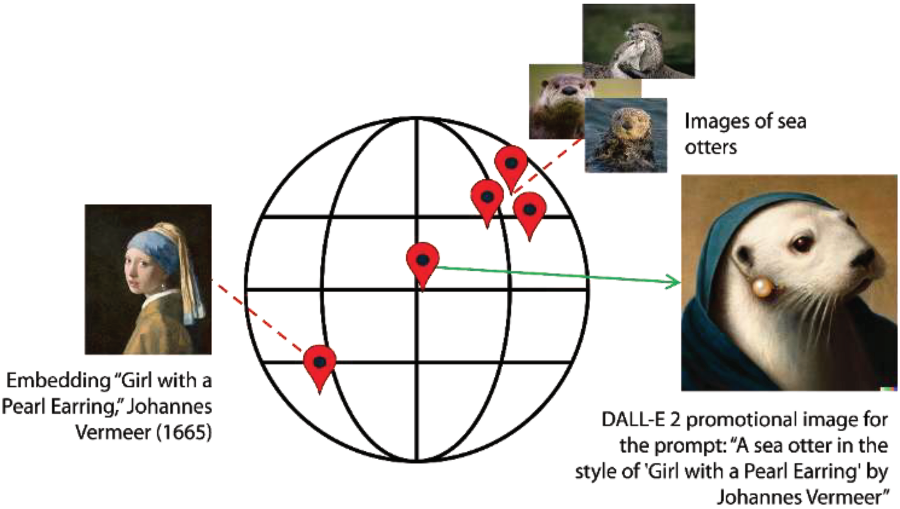

Figure 2: The embedding for an interpolated image, like this sea otter in the style of Vermeer, exists at a point in the latent space between the embedding for its inspirations, here the embeddings for Vermeer's painting and for otter pictures. When DALL-E 2 creates images like this, it does not load prior images into memory or interpolate them on the fly. Instead, it chooses a pre-existing embedding then processes it through a decoder to generate an image with features mathematically linked to the embedding's coordinates.

Although one could devise many methods for translating prompts to embeddings, DALL-E 2's method is noteworthy

109 *See* DALL-E 2 paper, *supra* note 3, at 6–7 (demonstrating how DALL-E 2 "blends" images by following a trajectory between their embeddings).

because it incorporates random sampling, yielding greater variety and reduced odds of reproducing specific training images. OpenAI achieves this translation through the awkwardly named "prior model," which it trained via a machine learning process using 250 million image-caption pairs, along with their CLIP embeddings, to devise a mathematical function that (1) matches any text embedding to a latent-space region,[110] and (2) selects a random embedding from that region.[111] The selection typically does not correspond to the embedding for any of the 250 million items used in training, but instead maps to one of the effectively infinite embeddings between or beyond them.[112]

One may wonder, why choose the image embedding at random from the region rather than taking the single best fit? One reason is to enhance output variety.[113] A model could be designed to choose the paradigmatic best match for each prompt. Perhaps it would learn, from the training data, that the paradigmatic cat is an orange, short-haired, adult tabby. It would respond to any cat prompt by depicting that specific cat unless the user specified otherwise (*e.g.*, "black cat" or "kitten"). By selecting a region and then randomizing, DALL-E 2 can produce responsive images but leave room for the cat to be young or old, long- or short-haired, and different in coloration or markings. The method trades consistency for versatility. The added

---

[110] The prior model does not receive the prompt directly; CLIP first translates the prompt into a text embedding. *See id.* at 5; *see also supra* note 96 (describing CLIP's use of text embeddings). This speaks to the models' interdependence in the system.

[111] For each instance of image generation, the prior model samples two image embeddings at random, then enlists CLIP to select the best match between the two. DALL-E 2 paper, *supra* note 3, at 5.

[112] Early DALL-E 2 users were surprised to find that nonsense words generated consistent pictures, including "Apoploe vesrreaitais" for birds and "Contarra ccetnxniams luryca tanniounons" for bugs or pests. Aaron J. Snoswell, *Did an AI Really Invent Its Own 'Secret Language'? Here's What We Know*, SCI. ALERT (June 7, 2022), https://www.sciencealert.com/did-an-ai-really-invent-its-own-secret-language-here-s-what-we-know [https://perma.cc/F786-Q7XR]. Lacking awareness of the underlying models, users speculated they had discovered the system's "secret language." *See id.* What they did not understand is that DALL-E 2's secret language is that of CLIP embeddings. The nonsense words generated consistent images because the prior model mapped the words to a consistent latent-space region. Just as the prior model reliably translates the word "poplar" to a latent-space region filled with trees, it reliably translates "Apoploe vesrreaitais" to a specific region that happens to correspond to birds. Apart from randomness, the likeliest explanation is that the nonsense words vaguely evoke the Latin naming scheme for animal species. *See supra* note 94 and accompanying text.

[113] *See* DALL-E 2 paper, *supra* note 3, at 12–13.

distance between the result and the user's precise request may also mitigate the likelihood of copyright infringement.[114]

3. *Image Diffusion*

Compared to visualizing the latent space, the process of creating an image through diffusion is straightforward. Although the mathematics are daunting, the upshot is that a decoder model takes a random pattern of noise and transforms it, one step at a time, until it becomes a clear image.[115]

Diffusion proper is the process of *adding* noise to an image.[116] At low enough noise levels, noisy images are recognizable. Once enough noise is added, however, noisy images may be virtually indistinguishable from television static:

Figure 3: Forward Diffusion (From Image to Noise)

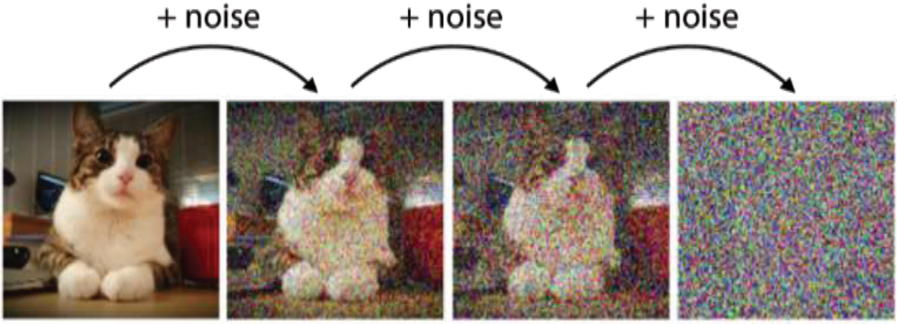

Figure 3: Diffusion incrementally transforms a cat picture into undecipherable static. At low noise levels, the cat is recognizable. At high noise levels, it is not. Images adapted from an NVIDIA research publication by Xiao, Kreis, and Vahdat.[117]

---

114 *See infra* section I.D.1.

115 *See* Kevin Schaul, Hamza Shaban, Shelly Tan, Monique Woo & Nitasha Tiku, *AI Can Now Create Images Out of Thin Air. See How It Works.*, Wash. Post (Dec. 17, 2022), https://www.washingtonpost.com/technology/interactive/2022/ai-image-generator [https://perma.cc/VE9T-NGF4].

116 Jascha Sohl-Dickstein, Eric A. Weiss, Niru Maheswaranathan & Surya Ganguli, *Deep Unsupervised Learning Using Nonequilibrium Thermodynamics*, arXiv 3–4 (Nov. 18, 2015), https://arxiv.org/pdf/1503.03585 [https://perma.cc/44F4-77EK] (describing the "forward trajectory").

117 Zhisheng Xiao, Karsten Kreis & Arash Vahdat, *Tackling the Generative Learning Trilemma with Denoising Diffusion GANs*, Nvidia (Mar. 2, 2022), https://research.nvidia.com/publication/2022-03_tackling-generative-learning-trilemma-denoising-diffusion-gans-0 [https://perma.cc/NL2P-3K43].

Diffusion-based image models train to conduct the reverse process of *subtracting* noise.[118] By analyzing images at different noise levels, the model learns to remove static and thereby make noisy images look incrementally more like the originals:[119]

Figure 4: Reverse Diffusion (From Noise Back to Image)

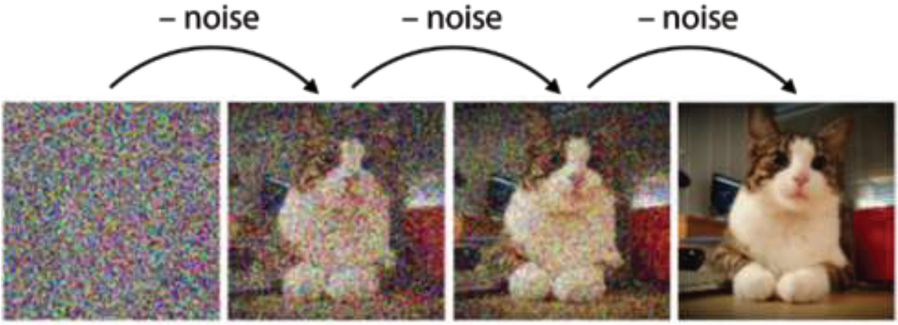

Figure 4: Reverse-diffusion models typically do not jump straight from noise to a clear, final picture. Instead, they remove noise one step at a time. The specific form this operation takes is a "Markov chain,"[120] meaning the model at each step needs only to know the immediately prior image (along with the embedding being expressed); the denoising process pays no attention to prior images in the chain. Images adapted from an NVIDIA research publication by Xiao et al.[121]

But these capabilities are not limited to pictures corrupted by noise. Once the model attains the ability to convert noise to an image with specific features, it can "restore" random noise patterns that were never pictures at all.[122] Each possible starting pattern is called a "seed."[123] Using random seeds adds variety: a model trained to reconstruct a picture of a particular

---

118 *See* Prafulla Dhariwal & Alex Nichol, *Diffusion Models Beat GANs on Image Synthesis*, ARXIV 3 (June 1, 2021), https://arxiv.org/pdf/2105.05233 [https://perma.cc/L4A7-Q9EL]. The technical term for subtracting noise is "annealing."

119 Nicholas Carlini et al., *Extracting Training Data from Diffusion Models*, ARXIV 2 (Jan. 30, 2023), https://arxiv.org/pdf/2301.13188 [https://perma.cc/8WGM-9FRK].

120 *See* Calvin Luo, Understanding Diffusion Models: A Unified Perspective at 5–6, arXiv (Aug. 26, 2022), https://arxiv.org/abs/2208.11970 [https://perma.cc/3F9S-LUML].

121 Xiao et al., *supra* note 117.

122 *See* Carlini et al., *supra* note 119, at 2; Schaul, Shaban, Tan, Woo & Tiku, *supra* note 115.

123 John Wolfe Compton, *How Seed Numbers Influence AI Image Generation*, https://johnwolfecompton.com/the-seed-of-imagination-how-seed-numbers-influence-ai-image-generation [https://perma.cc/BGP3-VC93] (last accessed Oct. 26, 2024).

breed of dog may construct different pictures for different starting seeds depending on where the pattern of static suggests the presence of a dog. To illustrate, consider the progression below. Because the patterns in true random noise are often too subtle for human perception, I inserted doctored seeds at Timestep 1 to show how a system would treat more obvious patterns when creating images for the prompt "dachshund snuggling under a blanket." The starting seeds are identical except that one includes a concentration of dog-shaped black pixels in the bottom left corner and the other instead concentrates dog-shaped brown pixels in the center. The diffusion process treats the concentrated pixels as the likeliest spot for the main figure and transforms the image accordingly:

Figure 5: Starting Noise Impacts Image Generation

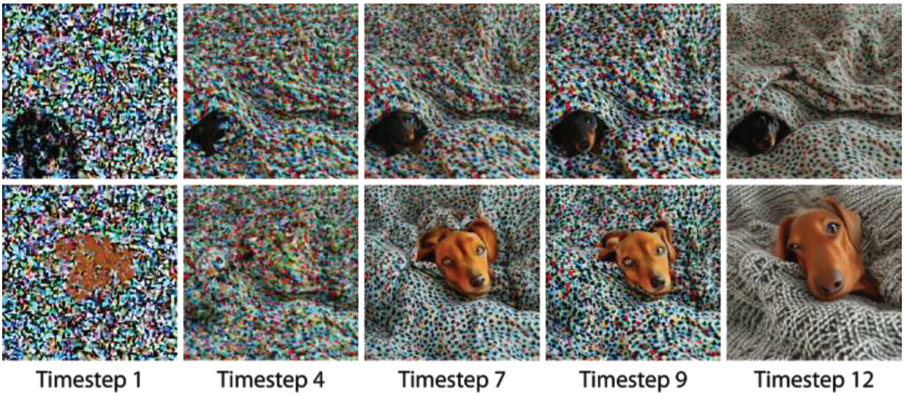

Figure 5: Diffusion models identify patterns in noise and mold those patterns to achieve the requested image, as illustrated through the progression above. Images produced by author using Stable Diffusion's "img2img" feature. All settings identical except variation of the Timestep 1 image.

OpenAI trained a decoder model to perform this reverse-diffusion task.[124] The decoder trained to remove noise using the same 250 million images and CLIP embeddings as the prior model.[125] This meant feeding pictures with various degrees of noise to the computer so it could devise mathematical processes

---

[124] Two final processing steps follow after the decoder creates an image. To conserve processing power, the decoder generates a small image at a base resolution of 64x64. DALL-E 2 paper, *supra* note 3, at 4. DALL-E 2 then employs two upsampler models—one to increase the image to 256x256 resolution, and another to generate the 1024x1024 resolution image delivered to the user. *Id.* The decoder's in-built 64x64 limitation may nonetheless bear on its ability to accurately reproduce training images. *Id.* at 17; *see infra* section I.D.2.

[125] *See* DALL-E 2 paper, *supra* note 3, at 23.

to reverse the noise and thereby reconstruct the original pictures.[126] Having learned these mathematical processes, the decoder model could apply them to different starting seeds for variations on the image.[127] We will return to the variations feature below to probe DALL-E 2's outputs.[128]

As with DALL-E 2's other systems, the objective was not merely to reconstruct or produce variations on the 250 million training images. OpenAI sought instead to devise generalizable mathematical functions that could yield never-before-seen images by decoding embeddings that did not correspond to training images.[129] This meant training the model to interpolate and extrapolate.[130] Recall that each embedding is effectively a set of latent-space coordinates in 319 dimensions. The model learned to assign image features depending on the coordinates it received. Given coordinates corresponding to training image *X*, it would reproduce features of that image; given coordinates for training image *Y*, it would reproduce features of that image instead. After training to perform this task for 250 million reference points, the resulting model could produce images for coordinates between and around those reference points. In effect, the decoder model learned how to treat every possible embedding as a set of instructions for decoding images with specific features.

In contrast to the other models, the decoder is DALL-E 2's only component that associates embedding coordinates to visual features of training images. Recall that CLIP was originally designed only to detect whether a given image embedding matches a given text embedding; although CLIP establishes the coordinate system for DALL-E 2's latent space, it cannot process images beyond mapping them to embeddings.[131] In the workflow for DALL-E 2, CLIP renders the user's prompt into a text embedding and sends that embedding to the prior model.[132] The prior model serves only to convert that text

126 *See id.* at 6, 23. The training and function of the DALL-E 2 decoder is similar to that of the prior OpenAI system "GLIDE." *See generally* GLIDE Paper, *supra* note 49. Given that OpenAI had also previously created CLIP, *see generally* CLIP Paper, *supra* note 49, the only wholly new feature of DALL-E 2 is the prior model.

127 *See* DALL-E 2 paper, *supra* note 3, at 6–8, 23.

128 *See infra* subpart I.D.

129 *See* DALL-E 2 paper, *supra* note 3, at 1–3.

130 *See id.* at 6–7.

131 *See supra* section I.C.1.

132 DALL-E 2 paper, *supra* note 3, at 3.

embeddings into image embeddings.[133] Given a prompt like "wooden Christmas tree ornament of Abraham Lincoln," it need not (and cannot) do anything except identify the corresponding latent-space region and select an embedding. The decoder, for its part, cannot decipher the user's prompt. But it takes the prior model's image embedding and seeks to find whatever image features it associates with those coordinates within a noise pattern. Diffusing these features allows the model to reconstruct the likeness of our sixteenth president in a small wooden object:

Figure 6: Decoding an Idiosyncratic Prompt

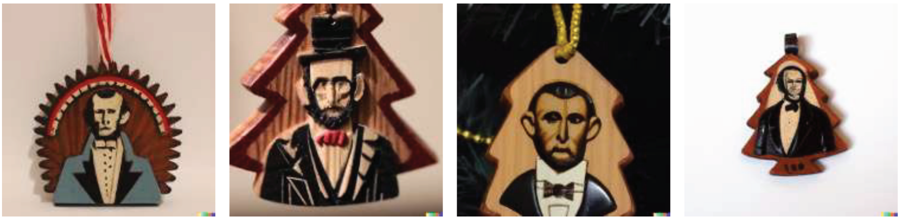

Figure 6: These images are DALL-E 2's first four outputs for the prompt "wooden christmas tree ornament of Abraham Lincoln." They are not cherry-picked. The rightmost image resembles Stephen Douglas; arguably, the leftmost does too. Mistakes like these provide insight into DALL-E 2's latent space. The oddity may arise because CLIP originally encoded images of the Lincoln–Douglas debate to a region near or overlapping with the region for Lincoln generally, resulting in ambiguity when the decoder model later trained on images mapped to embeddings in that region. Alternately, the oddity may arise from randomization in the prior model's embedding selection. Images produced by author using DALL-E 2.

D. Probing the Latent Space

Model training is just half the story. Many infringement and fair-use questions depend not only on what the models contain, but also on what one can do with them. The *Google Books* decision[134] exemplifies the point.[135] There was no question that Google had scanned and retained the complete text

---

133 *See supra* section I.C.2.

134 Authors Guild v. Google (*Google Books*), 804 F.3d 202 (2d Cir. 2015).

135 *See infra* section II.B.1.

for millions of copyrighted books.[136] Google established fair use not by disputing that it made copies, but by demonstrating that its system did not divulge the copies—it provided only limited access during normal operation and adopted safeguards against extraction.[137] Notwithstanding the likelihood that AI models likely memorize some training works to some degree,[138] fair-use analysis will require scrutinizing what generative systems can produce and what safeguards are warranted.

1. *Training, System Design, and Substantial Similarity*

Determining what is "in" a generative system is tantalizing. It is tempting to imagine that, if only we knew what each model retained from its training works, we would hold answers to several pressing copyright questions.[139] Answers remain elusive, however, because AI models save information in ways humans cannot directly scrutinize.[140] We are left to infer models' contents from their outputs.[141]

A generative system's outputs are the culmination of decisions made during model training and design of the final system. Although law and policy discussions recognize this point to a degree, they often focus on the extremes of deduplicating the training set at the beginning of training and using prompt filters or content filters during operation.[142] These decisions have obvious implications: studies demonstrate that training on duplicate images increases the likelihood of memorizing

---

136 *See* 804 F.3d at 207.

137 *See id.* at 209–10 (explaining the limits of Google's "snippet view"); *id.* at 227–28 (addressing hacking risks). *Google Books* did not use the "extraction" terminology; it instead spoke of "piratical hacking" in the event an attacker obtained access to Google's database. *Id.* at 228. I use the term to highlight the parallels between the portion of *Google Books* discussing unauthorized access and the ongoing discussion around extraction of training data from AI models. *See generally* Cooper & Grimmelmann, *supra* note 108.

138 *See* Cooper & Grimmelmann, *supra* note 108, at 50 ("some amount of memorization might even be *required* for effective generalization"). In DALL-E 2's case, this is particularly true if the decoder achieved the objective of reconstructing some number of training images given the correct starting seed. *See supra* note 126 and accompanying text.

139 *But see* Cooper & Grimmelmann, *supra* note 108, at 6–7 (resisting the claim that factual memorization compels specific legal conclusions).

140 DALL-E 2 paper, *supra* note 3, at 8–9.

141 *Cf.* 5 NIMMER & NIMMER, *supra* note 9, at § 20.05 (exploring technical and metaphysical difficulties of determining copies' existence within a model).

142 *See* Ohm, *supra* note 66, at 231. Ohm's work takes a different tack, emphasizing the benefits of intervening at the fine-tuning stage. *Id.*; *see also supra* subpart I.A (distinguishing training and fine-tuning).

those images.[143] Meanwhile, filters can thwart users' ability to request particular outputs.[144] These strategies reduce the likelihood that a system will regurgitate an image identical or substantially similar to a training image.[145]

Other training and design decisions may be just as impactful yet driven by concerns apart from copyright. Consider the difference between Stable Diffusion's and DALL-E 2's outputs for the prompt "Starry Night by Van Gogh":

Figure 7: "Starry Night by Van Gogh"

Figure 7: Stable Diffusion and DALL-E 2 images produced by author.

The original painting appears on the left. Stable Diffusion's results essentially duplicate the original, subject to cropping. DALL-E 2's results take looser inspiration. What accounts for the difference?

The answer may lie partly in curation of training images. Stable Diffusion may be afflicted by the aforementioned duplicates problem: it was trained on a dataset known for containing many duplicates and it would not be surprising for a dataset culled from the internet to contain several copies of a famous painting.[146] The duplicates would make memorization and subsequent regurgitation more likely.[147]

---

143 *See* Ryan Webster, Julien Rabin, Loic Simon & Frederic Jurie, *On the De-duplication of LAION-2B* at 1, ARXIV (Mar. 17, 2023), https://arxiv.org/pdf/2303.12733 [https://perma.cc/9M6T-N2HQ]; *see also* Carlini et al., *supra* note 119.

144 *See* Ohm, *supra* note 66, at 230–31.

145 Copyright infringement triggers at the threshold of "substantial similarity." *See infra* notes 160–61 and accompanying text.

146 *See* Webster, Rabin, Simon & Jurie, *supra* note 143, at 1 (calculating one third of LAION-2B's images as duplicates).

147 *See id.* at 2.

But DALL-E 2 may have retained more information about "Starry Night" than Figure 7 suggests. Recall that every DALL-E 2 image traces back to a specific point in its latent space.[148] DALL-E 2 allows us to navigate to specific points with precision through "image variations," a feature whereby the system encodes an uploaded image directly to an embedding and then generates new images from the corresponding point in the latent space.[149] Feeding "Starry Night" into the system in this manner yields the following:

Figure 8: DALL-E 2 Variations on "Starry Night"

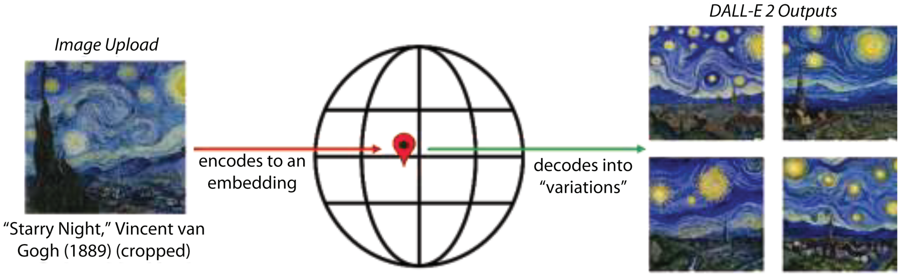

Figure 8: DALL-E 2 images on the right produced by author as variations of van Gogh's image on the left.

From this result, we can see that DALL-E 2 has the capacity to produce an image much closer to the original, and we can infer that the model retains several details from the original.[150]

The contrast between DALL-E 2's outputs in Figure 7 and Figure 8 points to the importance of training and design decisions beyond curation of training images. Stable Diffusion translates user prompts into latent-space embeddings in a deterministic fashion.[151] This means that typing "Starry Night by Van Gogh" into Stable Diffusion yields the same embedding every time.[152] If that point corresponds to detailed instructions for

148 *See supra* subpart I.C.

149 *See* DALL-E 2 paper, *supra* note 3, at 3.

150 The inference is not ironclad. Section I.D.2, *infra*, demonstrates the system's ability to reproduce the features of images not included in the training set.

151 *See* Yuxuan Ding, Chunna Tian, Haoxuan Ding, & Lingqiao Liu, *The CLIP Model Is Secretly an Image-to-Prompt Converter*, ARXIV 3 (Feb. 15, 2024), https://arxiv.org/pdf/2305.12716v2 [https://perma.cc/3YVG-3QQ3].

152 The Stable Diffusion user can nonetheless reduce this determinism by adjusting the "classifier-free guidance scale" at the time of operation, which essentially entails a decision of how heavily to weigh the visual called for by the embedding versus how heavily to weigh patterns discernible in the starting seed. *See* Chris McCormick, *Classifier-Free Guidance (CFG) Scale*, MCCORMICKML (Feb. 20,

reproducing the original painting—as indicated by Figure 7—then the system will yield that painting every time a user enters that prompt. By contrast, DALL-E 2 chooses embeddings more randomly via its prior model.[153] Running the prompt "Starry Night by Van Gogh" through the prior model does not select the most perfect fit for the prompt, but instead selects an embedding somewhere in the vicinity.[154] Although OpenAI plausibly incorporated randomness into the prior model to facilitate image diversity,[155] randomness may also steer the system away from replicating training images.

Other design decisions are more subtle still. Even assuming an AI model contained all information necessary to reproduce a training image, it might be configured in a manner that made reproduction impossible. Without the ability to study the models directly, the impossibility of the reproduction would preclude us from ever finding out that the information resided in the model.[156] OpenAI's own attempt to replicate an image illustrates the point:

Figure 9: Limits of Low-Resolution Decoding

Figure 9: Left, original image; right, DALL-E 2 decoding of the original's embedding. Image adapted from DALL-E 2 paper.[157]

2023), https://mccormickml.com/2023/02/20/classifier-free-guidance-scale [https://perma.cc/G5K7-53WD].

153 *See supra* section I.C.2.

154 *See id.*

155 *See supra* notes 113–14 and accompanying text.

156 *See* 5 NIMMER & NIMMER, *supra* note 9, at § 20.05[C][1][b].

157 DALL-E 2 Paper, *supra* note 3, at 18.

On the left is a training image, and on the right is DALL-E 2's failed attempt to reconstruct it. The dog on the right lacks key details such as the lake and the overall image composition. OpenAI attributes the failed reconstruction to a decision it made in configuring the final system: the diffusion model is configured to produce initial images only at a 64x64 resolution, which leaves some details present in the training images too fine to reproduce.[158] OpenAI speculates the output might be a closer match if it reconfigured the diffusion model to produce higher-resolution images.[159]

2. *Resemblance Versus Factual Copying*

AI outputs can resemble pre-existing human works without copying those works. Many scholars argue that outputs often fall short of *legal copying*, meaning the outputs do not infringe copyright because they lack substantial similarity to the training works.[160] The training and design decisions surveyed above can make this similarity more or less likely.[161] The additional point here is that outputs may sometimes fall short of infringement due to the absence of *factual copying*. Outputs can resemble specific human works even though those works were not in the training set and therefore could not be copied during the system's training. As we will explore later, this also means systems can produce outputs that *compete* with specific human works even without copying them.[162] The upshot is that, in some cases, actionable copying may occur only during training.[163] This is why evaluating the legality of copying during training is crucial.[164]

---

158 *Id.*

159 *Id.* at 17.

160 The argument is that—aside from cases of memorization—many outputs reproduce at most the "style" of the training works and this precludes infringement because style is generally unprotectable. *See, e.g.*, Murray *supra* note 9, at 305; Sag, *supra* note 9, at 342–43. Others question that assertion, most notably Ben Sobel in cautioning that the slogan "you can't copyright style . . . risks providing AI users with a false sense of security *and* misinforming copyright holders about the extent of their legal rights." Benjamin L.W. Sobel, *Elements of Style: Copyright, Similarity, and Generative AI*, 38.1 HARV. J.L. & TECH. (forthcoming 2025) (manuscript at 28) (May 18, 2024 draft on file with author).

161 *See supra* section I.D.1.

162 *See infra* subpart IV.B.

163 *But see* Cooper & Grimmelmann, *supra* note 108, at 16 (insisting "all generative-AI models *memorize* some portion of their training data").

164 *See supra* note 9 and accompanying text.

Our odyssey into DALL-E 2's latent space helps us see how resemblance is possible without copying. Most AI developers regard regurgitation of training data as a bug, not a feature.[165] The ideal image-generation system instead generalizes to create new works.[166] A system with sophisticated generalization capabilities may nonetheless yield an output that resembles a pre-existing work even if the work was never copied during training.

DALL-E 2's image-variations feature allows us to test this proposition: uploading an image is equivalent to asking the CLIP model to identify the latent-space embedding that best corresponds to the image and then asking the diffusion model to visualize that embedding.[167] We saw this above with "Starry Night," where DALL-E 2 did a passable job reconstructing an image that is likely in the training set.[168] Remarkably, DALL-E 2 can also construct images that resemble works that likely fall outside the training set. Observe its production of image variations following upload of the award-winning student art piece titled "Radiance" from 2023:

Figure 10: Successful Variations on Image Outside Training Set

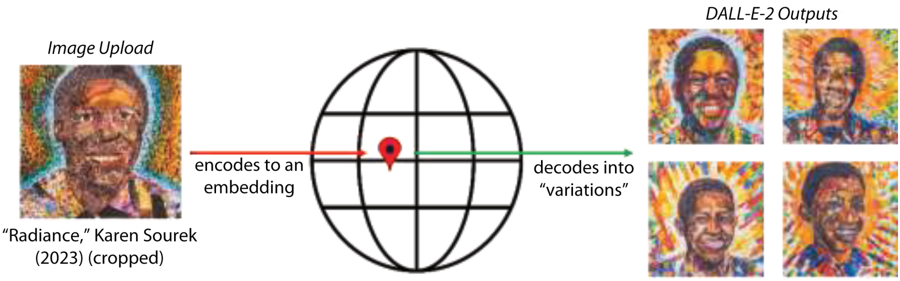

Figure 10: DALL-E 2 images on the right produced by author as variations of Karen Sourek's image on the left.

The result is remarkable because the outputs already existed as latent images within the system prior to my upload of "Radiance." The latent space constructed through the training

[165] *See* Cooper & Grimmelmann, *supra* note 108, at 63.

[166] Deven R. Desai & Mark Riedl, *Between Copyright and Computer Science: The Law and Ethics of Generative AI*, 22 Nw. J. Tech. & Intell. Prop. 55, 67 (2024) ("Generalization happens when the model learns patterns that can be applied to situations it has never encountered before.").

[167] *See supra* note 149 and accompanying text.

[168] *See supra* Figure 8.

of DALL-E 2's models contained an embedding that effectively embodied the image features of "Radiance"—the features on display in the four outputs above—by building off images available prior to DALL-E 2's launch in 2022. The latent image features only became apparent, however, when I navigated to the embedding. Uploading "Radiance" did not teach the system how to paint the picture, but instead gave the system the coordinates to access an embedding it had already interpolated.

To emphasize the point that the image existed in the prepopulated latent space, consider the following mistaken variation on another award-winning student piece titled "My Dog and I" from the same competition:

Figure 11: Mistaken Variations on Image Outside Training Set

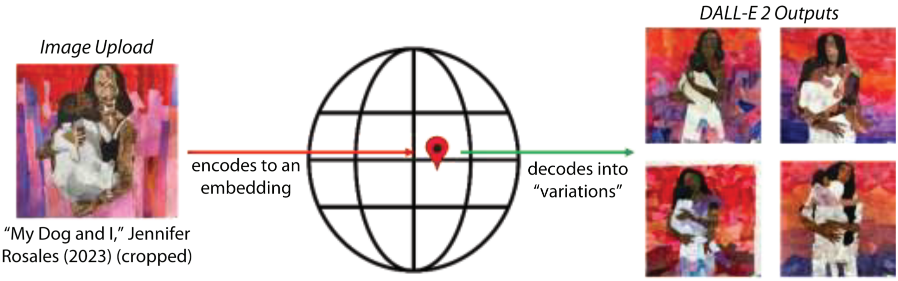

Figure 11: DALL-E 2 images on the right produced by author as variations of Jennifer Rosales's image on the left.

The original piece clearly features a woman holding a dog with a brown head and white body. The variations instead depict a woman holding a child with dark hair and white clothes. This mistake reminds us that the system's outputs are tied to whatever patterns the model took from the training set.[169] The result suggests that, in the training set, the typical picture of a woman holding a loved one features a mother and child. Because the training did not encompass enough contrary examples where a woman instead held her dog, the output reflects the more common scenario.[170]

169 *See supra* section I.C.3.

170 Race may also be a factor. White people are overrepresented in DALL-E 2's training data relative to non-white people, *see DALL-E 2 Pre-Training Mitigations*, OPENAI (June 28, 2022), https://openai.com/research/dall-e-2-pre-training-mitigations [https://perma.cc/ZUY8-PMED], meaning that white people in the training data may be engaged in a more diverse array of behavior (including, presumably, holding dogs). The smaller number of reference images for Black

3. *The Spider-Man Problem*

Copyright's treatment of fictional characters extends the scope of substantial similarity for images involving copyrighted characters. Ordinarily, regurgitation results in an infringing reproduction[171] while generalization presents more subtle questions.[172] The generalized image may constitute an infringing reproduction if it hews too closely to the composition and other expressive choices evident in a training work,[173] but copyright leaves room to emulate style and subject matter without infringement.[174] That room is diminished for copyrighted characters, however, because courts have held that the copyright for a clearly delineated character can be infringed through any further depictions in which the character is recognizable.[175]

Enter the Spider-Man problem. Matthew Sag has previously discussed the "Snoopy Problem," which refers to AI models' propensity to retain and reproduce recognizable character features.[176] An AI model trained on 1,000 distinctive pictures labeled as "Snoopy" may avoid memorizing any one picture yet internalize the character's appearance as a cartoon beagle with a long, rounded snout who often dresses as a pilot and appears with a small yellow bird.[177] To his observation I add a further corollary I call the "Spider-Man problem"—the difficulty of stopping AI systems from reproducing famous characters without also trenching on legitimate expression.

Characters like Spider-Man are ubiquitous in popular culture and their popularity all but guarantees that image sets taken from the internet at large or user uploads will contain several representative depictions. It is evident from DALL-E 2's responses to prompts asking for Spider-Man

and Latina women may leave the system less capable of depicting a broad range of activities for non-white women like the central figure in "My Dog and I." *See* Leonardo Nicoletti & Dina Bass, *Humans Are Biased. Generative AI Is Even Worse*, BLOOMBERG TECH. (June 9, 2023), https://www.bloomberg.com/graphics/2023-generative-ai-bias [https://perma.cc/7N9A-DU7W].

171 *See* 5 NIMMER & NIMMER, *supra* note 9, at § 20.05[C][2][a].

172 *See id.*

173 *See id.*

174 *See id.*

175 *See id.* at § 20.05[C][2][b].

176 Sag, *supra* note 9, at 327–34.

177 *Id.* at 334.

that training on such images has populated DALL-E 2's latent space with countless images and permutations upon the character:

Figure 12: Reproducing Fictional Characters

"The Amazing Spider-Man" | "Peter Parker unmasked" | "Andy Warhol in Spider-Man costume" | "Pikachu in a Spider-Man costume in the style of Van Gogh"

Figure 12: These images are not cherry-picked—each comes from DALL-E 2's first batch for the listed prompt. The results demonstrate not only the system's ability to faithfully depict Spider-Man, but also its association of Peter Parker with Spider-Man and ability to interpolate Spider-Man's likeness with real figures (Andy Warhol), other fictional characters (Pikachu), and art styles (Van Gogh's). DALL-E 2 images produced by author.

In further testing, I found that Meta AI produced a recognizable Spider-Man in response to the incomplete prompt "Draw an acrobatic superhero in". This result speaks to the pervasiveness of Spider-Man even in Meta's training set, which was purportedly limited to user uploads:[178]

---

178 *See* Edwards, *supra* note 4.

Figure 13: Spider-Man Interrupted?

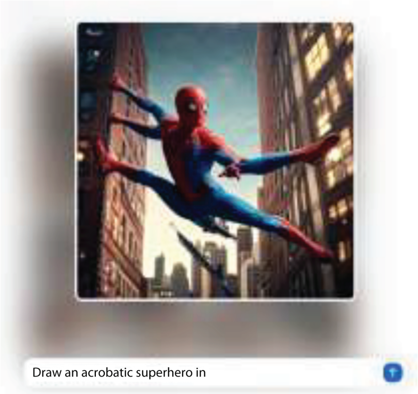

Figure 13: Meta AI provides an "imagine" feature that allows real-time visualization of images as the user adds words to the prompt. Real-time visualization of the fragmentary prompt "Draw an acrobatic superhero in" was sufficient to invoke a recognizable Spider-Man. Meta AI image produced by author.

Later image systems like DALL-E 3 and Google's Gemini have attempted to combat these results by refusing to run prompts that expressly mention copyrighted characters' names. Because these interventions consist only of filtering at the time of prompting, however, they do not eliminate the latent-space regions associated with Spider-Man.[179] The system accesses this region in response to prompts like "draw a superhero in a

[179] *Cf.* Ohm, *supra* note 66, at 231 ("[I]nput and output filters do not change the underlying model.").

red and blue costume with webbed pattern swinging through New York City on a white rope" with predictable results:

Figure 14: Reproducing Fictional Characters Without Names

DALL-E 3 Google Gemini

Figure 14: Both systems rejected direct requests to generate "Spider-Man." The DALL-E 3 output came in response to the prompt above. The Gemini outputs respond to "Generate a superhero in a tight-fitting red and blue costume, with web pattern and large, white eyes, swinging across New York City while holding a long strand of webbing." DALL-E 3 and Gemini images produced by author.

Any attempt to avert these results would be costly and could impose collateral consequences on user expression. The solution is not so simple as merely removing literal duplicates. Prominent characters like Spider-Man make cameo appearances in patents,[180] Halloween photos,[181] and coverage

---

180 *See* Kimble v. Marvel Ent., LLC, 576 U.S. 446, (2015) (discussing a web-shooter toy patented under U.S. Patent No. 5,072,856 (filed May 25, 1990)).

181 *See* Kaity Kline, *It's a Pink Halloween. Here Are Some of the Most Popular Costumes of 2023*, NPR (Oct. 18, 2023), https://www.npr.org/2023/10/18/1206356916/its-a-pink-halloween-here-are-some-of-the-most-popular-costumes-of-2023 [https://perma.cc/WJ4H-MHTT] (projecting "2.6 million planning to dress as the superhero").

of street performers in Las Vegas and Times Square,[182] to say nothing of films and innumerable tie-in products depicting the character on advertisements and packaging.[183] Filtering all such images would be costly and could screen out much more than just Spider-Man. Katrina Geddes warns that the result would be to undermine semiotic democracy by removing users' ability to create critical, countercultural, or playful character images that advance copyright's objectives and plausibly qualify as non-infringing fair uses.[184] More mundanely, excising these images could undermine copyright's expressive values by stripping the verisimilitude of outputs meant to show children on Halloween or accurately depict Times Square.

There is also room to question whether these costs would be justified relative to their effectiveness. It may take only a handful of stray images—children's superhero-themed birthday cakes,[185] or the Spider-Man float from Macy's Thanksgiving Day Parade[186]—for a model to chart a latent-space region for Spider-Man. Indeed, current-generation systems can generalize from text to produce a recognizable Spider-Man using a model trained on exactly zero images of Spider-Man. Observe

182 *See* Tim Kenneally, *Fake Spider-Man Gets in Real Brawl in Front of Times Square Toys 'R' Us (Video)*, THE WRAP (Aug. 7, 2015), https://www.thewrap.com/fake-spider-man-gets-in-real-brawl-in-front-of-toys-r-us-video [https://perma.cc/8J2G-LCF9].

183 See 17 U.S.C. § 113(c), which makes it perfectly lawful to photograph many products depicting Spider-Man and circulate the images widely "in connection with advertisements or commentaries related to the distribution or display of such articles."

184 *See* Geddes, *supra* note 40, at 34–38.

185 *See* Anita Butterworth, *19 Spiderman Cake Ideas for Super Birthdays*, MOUTHS OF MUMS (Mar. 22, 2023), https://mouthsofmums.com.au/spiderman-cake-ideas [https://perma.cc/45TG-Y2FG].

186 *See* Daniel Dockery, *Thanksgiving is Spider-Man's Holiday*, POLYGON (Nov. 23, 2023), https://www.polygon.com/23971196/spider-man-thanksgiving-connection-holiday [https://perma.cc/FXY5-92RT].

the results of asking ChatGPT for the code to draw Spider-Man in a format called TikZ:

Figure 15: ChatGPT Draws Spider-Man

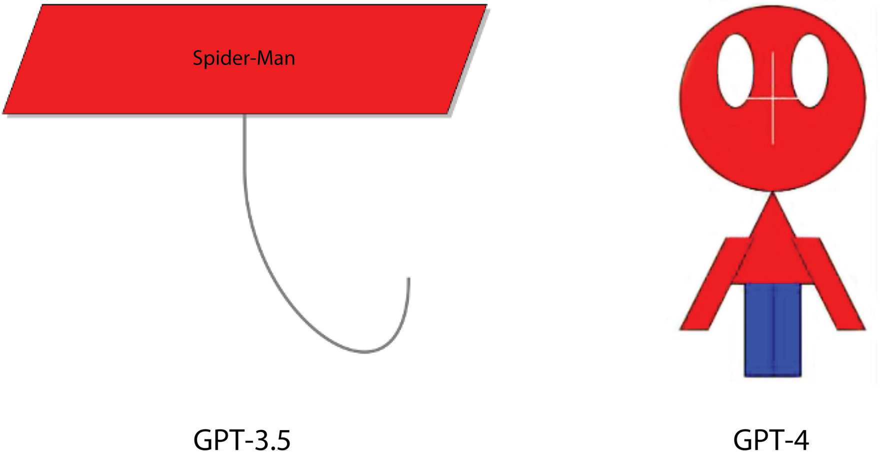

Figure 15: LaTeX images produced by author using ChatGPT's textual responses to the prompt "Produce code to create a TikZ drawing of Spider-Man" using GPT-3.5 (left, note the trailing "web") and GPT-4 (right). TikZ is a coding language used for creating basic images using the document-typesetting program LaTeX.

Although both drawings are rudimentary, they are noteworthy outputs for AI models trained without images. The GPT models that power the ChatGPT system train exclusively on text,[187] meaning their training set included no Spider-Man images because it included no images at all. Yet the model evidently extrapolated Spider-Man's appearance from text alone.

There is sure to be considerable debate regarding how to handle the Spider-Man problem, particularly with respect to what steps must be taken to avoid secondary liability. Efforts to cabin infringing outputs may also be pertinent to assessing fair use. The present discussion seeks not to resolve the debate, but to emphasize the difficulty of stamping out such outputs and the tradeoffs for public expression.

* * *

187 *See* Sébastien Bubeck et al., *Sparks of Artificial General Intelligence: Early Experiments with GPT-4*, ARXIV 8 (Apr. 23, 2023) https://arxiv.org/pdf/2303.12712 [https://perma.cc/8CKM-KT7A] ("This demonstrates that GPT-4 can 'see' despite being a pure language model . . . .").

The foregoing technical introduction sets the stage for the fair-use analysis. We could have assumed many pertinent features—for example, that model trainers and system designers sometimes operate independently, that training sometimes pursues narrow technical objectives distinct from those of the assembled system, and that it is possible to create models and systems whose outputs do not meaningfully replicate training images. What our deep dive adds is context to understand why AI systems are constructed in this fashion and how they might be built differently. In doing so, it also provides the scaffolding for future work examining copyrightability and theories of liability beyond the scope of the present fair-use discussion. As to the present matter, this context prepares us to confront those features of generative AI that diverge from the standard assumptions underlying fair use.

II
Three Challenges for Fair Use

Generative AI presents deep questions for fair use. But the questions are not new. Nor are they intrinsic to the technology. Instead, AI has brought fresh urgency to uncertainties already present in fair-use doctrine and its normative underpinnings. The conventional account of transformative use requires evaluating the use's purpose and implicates questions around the scope of the relevant use—do we assess the purpose of copying during training by reference to the later system or do we evaluate training in isolation? Recent Supreme Court guidance focusing on "substitutability" draws attention to potential differences among kinds of substitutionary harm—how do we weigh substitution that follows from copying facts or other uncopyrightable elements? Scholars have proposed that non-expressive uses are fair, broaching the question of what it means for a use to be non-expressive. Is it sufficient that the system refrains from creating actionable copies, or must the training process be indifferent to the training works' expressive content?

These questions lack clear answers because, until now, they have been peripheral in most cases. Most prior cases could be decided the same under any plausible answer within any of these paradigms. The following discussion will detail how generative art makes these questions central and potentially dispositive. It begins by introducing transformative use.[188] It then works through three competing doctrinal approaches to transformativeness that I dub the entanglement approach, disaggregation approach, and substitutability test to explain why each is lacking.[189] It closes by addressing theories of non-expressive use.[190] The floundering of transformative use in the face of generative AI provides an opening for non-expressive use to carry distinct explanatory power. To meet that challenge, however, these theories too require closer scrutiny.

A. Transformative Purpose

Fair use is a complete defense to copyright infringement structured around four statutory factors: (1) purpose of the defendant's use, (2) nature of the plaintiff's work, (3) amount used, and (4) market impact.[191] Although Congress has recognized these factors statutorily, the statute provides no rubric for weighing them and the courts retain discretion to develop the relevant standards.[192]

Fair use has achieved a measure of predictability despite its open-endedness because courts have converged around the paradigm of transformative use.[193] Formally, transformative use is merely a sub-factor under factor one, where establishing a transformative purpose favors the defense.[194] In practice, however, it has become the *de facto* test because transformativeness shapes the remaining analysis.[195] Under factor four, for example, transformativeness may suggest market harm is unlikely because the defendant operates in a market distinct from any the plaintiff would ordinarily enter.[196]

---

188 *See infra* subpart II.A.

189 *See infra* subparts II.B–II.C; *see also* 5 NIMMER & NIMMER, *supra* note 9, at § 20.05[D][2] (delineating the three approaches and their shortcomings).

190 *See infra* subpart II.D.

191 17 U.S.C. § 107.

192 *See* 4 NIMMER & NIMMER, *supra* note 9, at § 13F.03[A].

193 *See* Clark D. Asay, Arielle Sloan & Dean Sobczak, *Is Transformative Use Eating the World?*, 61 B.C. L. REV. 905, 942 (2020); Barton Beebe, *An Empirical Study of U.S. Copyright Fair Use Opinions Updated, 1978–2019*, 10 N.Y.U. J. INTELL. PROP. & ENT. L. 1, 25 (2020); Jiarui Liu, *An Empirical Study of Transformative Use in Copyright Law*, 22 STAN. TECH. L. REV. 163, 180 (2019).

194 *See* Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 578–79 (1994).

195 *See* 4 NIMMER & NIMMER, *supra* note 9, at § 13F.10[B][2]; *see also* sources cited *supra* note 193.

196 *Campbell*, 510 U.S. at 590–91.

From 2006 onward,[197] the prevailing approach for determining whether a use is transformative has been to ask whether the use possesses a transformative purpose relative to the work copied.[198] It is not enough to add new context and material. Take courts' repeated rejections of fair use even for elaborate reworkings of television material, reasoning that the follow-on materials overlap with the shows because they seek to entertain by telling the same stories.[199] By contrast, courts have recognized the use of entertainment or promotional materials as transformative when the defendant has recontextualized the materials in documentary contexts.[200] Of special relevance to generative AI, courts have upheld many technological uses as fair despite literal copying of vast numbers of works.[201] As two noteworthy image search-engine cases explained, the informational purposes of copying images to create a search engine diverge almost entirely from the aesthetic purposes behind the original images.[202]

Although transformative use has drawn many critics and proposed alternatives—including theories of non-expressive

---

197 *See* 4 NIMMER & NIMMER, *supra* note 9, at § 13F.10[C] (identifying a sea change in transformative use).

198 Andy Warhol Found. v. Goldsmith, 598 U.S. 508, 542 (2023) ("[W]hether the new use served a purpose distinct from the original, or instead superseded its objects . . . was, and is, the 'central' question under the first factor.").

199 *See, e.g.*, Castle Rock Ent., Inc. v. Carol Publ'g Grp., Inc., 150 F.3d 132, 143 (2d Cir. 1998) (rejecting transformative purpose in the *Seinfeld Aptitude Test*); Paramount Pictures Corp. v. Carol Publ'g Grp., 11 F. Supp. 2d 329, 335 (S.D.N.Y. 1998) (reaching similar conclusions for *The Joy of Trek*), *aff'd*, 25 F. Supp. 2d 372 (2d Cir. 1999). Likewise, *Warner Bros. Ent. Inc. v. RDR Books*, 575 F. Supp. 2d 513 (S.D.N.Y. 2008), held that an unauthorized reference text on *Harry Potter* was "not consistently transformative" because it failed to minimize the "expressive value" of the expression taken, *id.* at 544, and another decision rejected transformativeness for incorporation of Abbot and Costello's "Who's On First?" routine into a theatrical production because the use owed its humor and impact to the original routine, TCA Television Corp. v. McCollum, 839 F.3d 168, 181–82 (2d Cir. 2016).

200 *See, e.g.*, Bill Graham Archives v. Dorling Kindersley Ltd., 448 F.3d 605 (2d. Cir. 2006). Defendants in this context may nonetheless face difficulty if they use too much: even though they may subjectively pursue a different purpose, substantial copies may inadvertently be capable of fulfilling the original work's purpose. *See, e.g.*, *Warner Bros. Ent.*, 575 F. Supp. 2d at 544.

201 *See, e.g.*, Kelly v. Arriba Soft Corp., 336 F.3d 811 (9th Cir. 2003) (image search engine); Perfect 10, Inc. v. Amazon.com, Inc., 508 F.3d 1146 (9th Cir. 2007) (same); A.V. *ex rel.* Vanderhye v. iParadigms, LLC, 562 F.3d 630 (4th Cir. 2009) (plagiarism detection); Authors Guild v. HathiTrust, 755 F.3d 87, 97 (2d Cir. 2014) (book digitization); Authors Guild v. Google Inc. (*Google Books*), 804 F.3d 202 (2d Cir. 2015) (same); *see also* 4 NIMMER & NIMMER, *supra* note 9, at § 13F.14[B] (tracing indexing and digitization decisions).

202 *See Kelly*, 336 F.3d at 811; *Perfect 10*, 508 F.3d at 1146.

use[203]—it has worked remarkably well for a range of new technology cases since the turn of the millennium. This is because identifying the purpose of the use has been tractable and purpose has been a workable proxy for substantive concerns including market impact and exploitation of the author's expression. Generative AI challenges this paradigm because it accentuates longstanding ambiguities in how we define a use's purposes and weigh its impacts.

B. Scope of Use

Generative AI exacerbates questions surrounding assessment of purpose for a multi-step copying process. The unspoken first step in identifying purpose is to decide the scope of the relevant conduct. One could evaluate the purpose of an act of copying by reference to the later uses it facilitates, which I dub entanglement, or one could evaluate the act of copying in isolation, which I dub disaggregation. The choice matters because the purpose of copying to train an AI model may, at some phases, be contingent and inchoate.[204] If we entangle a model's training process with the completed system and its outputs, we may conclude that the copying is non-transformative because it pursues expressive purposes that overlap with those of the training works. By contrast, if we disaggregate the training process from the final system, we may conclude that its purpose is transformative. Which vantage is correct? What criteria should drive the decision? And what of scenarios in which uncoordinated, unrelated parties engage in separate phases of creating an AI system?

It is not just that generative AI makes these questions complicated. It also makes them matter in ways they previously did not. Caselaw remains undeveloped because the decision has little impact in the typical case. Most prior cases that implicated intermediate copying, at some early stage, involved final products that also produced or contained copies. This meant that, even if a court considered preliminary or intermediate copying in isolation, the final product would face scrutiny for its independent acts of copying. Not so for generative AI. To the extent generative AI can be designed not to produce or retain copies at later steps, verifiable copying may occur only during training. Moreover, framing the choice between

---

203 *See infra* subpart II.D.

204 *See supra* subpart I.A.

the contrasting approaches as a binary leaves little room for nuanced analysis: focusing narrowly on technical purposes strongly favors fair use whereas considering the full system and its impacts weighs heavily against it. Recourse to prior reverse engineering decisions helps illustrate why the choice takes on outsized significance and suggests that purpose may be the wrong question here.

1. *Entanglement Approach*

Several leading new-technology cases follow the entanglement approach: they treat the purpose of intermediary copying involved in creating a system as coextensive with that of the completed system.[205] *Google Books* illustrates.[206] Authors challenged Google's wholesale scanning and digitization of millions of books to create a book search engine, leading to a fair-use decision. The *Google Books* court examined the service in which these scans were ultimately used, holding that the service's purpose was distinct from the books' and therefore transformative.[207] Whereas the books' authors wrote to convey expressive or informational content, Google's scans served the distinct purpose of creating a searchable index.[208] Although Google Books provided book snippets in response to search queries, the court deemed these snippets too short to usurp the books' expressive purposes.[209] The transformativeness of the service was underscored by its capability to facilitate other purposes beyond those of the original books, like Google's ngrams tool, which "allows readers to learn the frequency of

---

205 Lee, Copper & Grimmelmann *supra* note 23, at 110; Lemley & Casey, *supra* note 31, at 764. The point generalizes to low-tech cases involving new works. Indeed, some cases involving allegedly infringing movie scripts, for example, rejected theories of infringement based on intermediate copying in the making of the script and focused exclusively on the final version. Matthew Sag, *Orphan Works as Grist for the Data Mill*, 27 BERKELEY TECH. L.J. 1503, 1530–32 (2012). However, this approach is not universal. *See* Walker v. Univ. Books, Inc., 602 F.2d 859, 864 (9th Cir. 1979) (holding defendant liable for infringing blueprints).

206 Authors Guild v. Google Inc. (*Google Books)*, 804 F.3d 202 (2d Cir. 2015); *see also* Authors Guild v. HathiTrust, 755 F.3d 87 (2d Cir. 2014) (applying similar reasoning to educational book service).

207 804 F.3d at 216–17.

208 *Id.* *See also HathiTrust*, 755 F.3d at 97 ("There is no evidence that the Authors write with the purpose of enabling text searches of their books.").

209 *See Google Books*, 804 F.3d at 224–25; *see also infra* section II.C.2 (addressing marginal cases where Google Books might usurp markets for certain fact-seekers).

usage of selected words in the aggregate corpus of published books in different historical periods."[210]

The court concluded that Google's copying to create the service was fair because the service pursued a transformative purpose. It did not belabor the relation between the preliminary copying and the final product, nor did it scrutinize the purpose of the digitization process apart from the final system. When library participants took the digitized books and contributed them to the separate HathiTrust Digital Library Project,[211] the courts likewise focused their analyses on the purpose of HathiTrust's service.[212] They did not bog themselves down with separate assessments of the acts of copying and transmission necessary to create the service, nor did they confront the complications that might follow from acknowledging that Google's purposes at the time of digitization may not have subjectively included HathiTrust's later use.[213]

The entanglement approach fits awkwardly with most generative systems. Entanglement presumes linkage between preliminary copying and the ultimate system. This link may be lacking for generative systems due to their modularity. As noted above, AI models incorporated into an art system may be created by unrelated parties for a different purpose or no specific purpose at all.[214] The same CLIP model used to map the latent space for a generative art system might be used instead to create a gardening app that scans plant photographs to diagnose agricultural maladies.[215] To say that the purpose of training a model depends on the purpose of the later systems in which the model may be incorporated is to say that the purpose is contingent and indeed inchoate at the time of copying.[216]

The awkwardness may be ameliorated in scenarios where the party who creates the AI model is the one who later

---

210 804 F.3d at 217.

211 *HathiTrust*, 755 F.3d at 90, 92 n.3.

212 *See id.* at 97–104.

213 I thank Robert Brauneis for flagging this and related complications.

214 *See supra* subpart I.A.

215 *See, e.g.*, Wei, Chen & Yu, *supra* note 78, at 1 ("[C]ross-modal retrieval is achieved in the developed system, facilitated by a novel CLIP-based vision-language model that encodes both disease descriptions and disease images into the same latent space.").

216 As Jane Ginsburg puts it: "If the lawfulness of the inputs turns on the character of the outputs, one cannot determine either *ex ante*." Jane C. Ginsburg, *Fair Use in the US Redux: Reformed or Still Deformed?*, 2024 SING. J.L. STUDS. 52, 89.

incorporates it into a system. Further complications nonetheless arise when the system's outputs fall short of the substantial similarity threshold—meaning that no copying occurs at the time of operation.[217] Applying entanglement to such a system would mean defining the purpose of an earlier act of copying as coextensive with a later act of non-copying. This extension of the entanglement approach may feel tenuous, as it begins to resemble the causation-bending logic of quantum entanglement—what Albert Einstein famously dismissed as "spooky action at a distance."[218] Most prior entanglement cases did not present this issue because they involved at least two separate rounds of copying: copying to create the system and then the reconveyance of partial or lower-quality copies at the time of operation.[219] It would seem the absence of further copying should be material to whether the use is transformative or otherwise fair, but it is not clear that the entanglement approach recognizes this point.

Focusing on the ultimate use may support the conclusion that generative systems are non-transformative and therefore training is unfair.[220] Within the parameters of transformative use, the argument is that the system lacks transformativeness because its outputs serve aesthetic purposes like those of the training works.[221] This resolution is unsatisfying, however, insofar as it disregards the different decisions that could be made during training and deployment to mitigate memorization, regurgitation, and infringement risk more generally.[222]

2. *Disaggregation Approach*

Disaggregation would instead identify the purpose of each act of copying in isolation. Although the disaggregation approach has historically been less prominent, the Supreme

---

[217] *See supra* section I.D.1.

[218] *See* Chris Ferrie, *Quantum Entanglement Isn't All that Spooky After All*, SCI. AM. (Feb. 13, 2023), https://www.scientificamerican.com/article/quantum-entanglement-isnt-all-that-spooky-after-all [https://perma.cc/98V3-9DRE].

[219] In a non-technology case involving creation of a new work, *see supra* note 205, it may be that "copying" in the sense of the Section 106 reproduction right only occurs at the time of creation. *See* 17 U.S.C. § 106(1). Any later exploitations of the new work would nonetheless trigger separate Section 106 violations for, say, distribution or public performance in connection with the unauthorized copy. *See id.* § 106(3)–(6).

[220] *See* Sobel, *supra* note 9, at 80–81.

[221] The argument holds best for systems that train on works in a specific modality (such as images) to create works in the same modality.

[222] *See supra* subpart I.D.

Court's 2023 *Warhol* decision appears to mandate it.[223] Specifically, it holds that fair-use analysis should not focus on the infringing work *per se* but instead on the specific act alleged to infringe.[224]

The Court adopted this approach in the course of rejecting the argument that the alleged transformativeness inherent in Warhol's portrait based on a copyrighted photo, or any intent Warhol may have had when he created the portrait in 1984, was relevant to determining the purpose of the portrait's licensure for a magazine cover in 2016.[225] The Warhol Foundation argued that the original photographer sought to depict Prince as human and vulnerable, whereas Warhol's "flattened, cropped, exotically colored, and unnatural depiction" commented on "the dehumanizing culture of celebrity in America."[226] The Court characterized the linkage between the magazine usage and work's initial creation as "false equivalence" and focused only on the former.[227] Because Goldsmith had licensed celebrity photos for the same kinds of magazine uses—indeed, she licensed the specific Prince photo at issue for Warhol's own prior use as an artist's reference for an earlier magazine illustration—the Court concluded the 2016 use served the same purpose and therefore was non-transformative.[228]

To be sure, there is room to distinguish *Warhol* from AI training. One might argue that *Warhol* merited disaggregation because the acts of copying were separated by thirty-two years and the earlier user (Warhol) had no say in the later use (having died in 1987).[229] Cases like *Google Books* could be distinguished because they involved acts of copying that were close in time and part of a common scheme by a single company.[230] Perhaps, then, entanglement remains good law for systems like Google Books and for contemporary AI systems so long as model training and system deployment form a common scheme.

---

223 *See generally* Andy Warhol Found. v. Goldsmith, 598 U.S. 508 (2023); 4 Nimmer & Nimmer, *supra* note 9, at § 13F.10[G][2][a][ii].

224 598 U.S. at 549 ("focus[ing] on the specific use alleged to be infringing"); *id.* at 554 (Gorsuch, J., concurring).

225 *See id.* at 533–34 (majority).

226 *See id.* at 566 (Kagan, J., dissenting).

227 *Id.* at 534 n.10 (majority).

228 *Id.* at 534–38.

229 *See* Douglas C. McGill, *Andy Warhol, Pop Artist, Dies*, N.Y. Times (Feb. 23, 1987).

230 *See generally* Authors Guild v. Google Inc. (*Google Books*), 804 F.3d 202 (2d Cir. 2015).

However, scenarios where one party trains the AI model(s) and another configures the model(s) into a system may remain difficult to distinguish.[231] When separate AI models are created by separate parties, or with no specific system in mind, it becomes harder to justify departure from *Warhol*'s instructions to evaluate each act of copying on its own terms.

Disaggregation would strengthen the claim that training is fair. Viewed in isolation, many acts of copying during model training entail a transformative purpose because the copying goes toward non-expressive tasks like mapping a latent space[232] or calibrating diffusion algorithms.[233] Although the trained diffusion model may later be responsible for creating new works that resemble the training works, that result would be contingent on the model's later incorporation into a system providing this functionality.[234]

The disaggregation approach may nonetheless lack nuance because it suggests that copying during training is transformative, and presumably fair, regardless of the whether a model creator exercises care during model training or system design.[235] Concluding that all such copying is fair may be warranted if we determine that training is equivalent to extracting unprotectable elements from prior works,[236] but that premise requires testing in a richer framework than entanglement versus disaggregation.[237] Moreover, to the extent that courts and policymakers find across-the-board insulation unsatisfying, they may simply develop more byzantine liability theories—they could insist that models contain copies, making subsequent copying or transfer of the model actionable,[238] or they could contort secondary liability doctrine to reach key actors in the supply chain.[239]

---

231 *See supra* notes 214–15 and accompanying text.

232 *See supra* sections I.C.1–2.

233 *See supra* section I.C.3.

234 *See supra* subpart I.A.

235 Even if defensible in theory, the equities and optics of a test skewed categorically in favor of training might lead courts to reject it. *See* Lemley & Casey, *supra* note 31, at 746.

236 *See* Bracha, *supra* note 9, at 21 (explaining "the only purpose of the reproduction is the extraction of metadata necessary for the machine learning process"); Murray, *supra* note 9, at 283–84 (describing extraction and use of metadata during training).

237 *See infra* subpart II.D & Part III.

238 *See* 5 NIMMER & NIMMER, *supra* note 9, at § 20.05[C][1][b].

239 *See id.* at § 20.05[C][3][b].

3. *Beyond Transformative Purpose*

Prior cases dealing with reverse engineering confronted similar difficulties with purpose. Viewed from either the entanglement or disaggregation perspective, the purpose of the copying overlapped with that of the original.[240] Yet courts have repeatedly upheld these uses as fair and sometimes even transformative. Any coherent rationale must go deeper than purpose differentiation.

Sega v. Accolade[241] was the first decision to uphold reverse engineering as fair.[242] Accolade sought to release games for Sega's popular Genesis console, but Sega had programmed the Genesis only to launch games that incorporated its proprietary lock-out code.[243] Accolade overcame the obstacle through a reverse engineering process that involved copying an authorized Genesis game multiple times to isolate the lock-out code; it then used that code to release games that competed with Sega's.[244] The Ninth Circuit upheld the copying as fair because it was the only practical way to obtain the lock-out code, which was excluded from copyright protection due to its functionality.[245] Because the case was decided before transformative use was dominant, the court did not address that theory.[246]

The next major decision also dealt with video games. In *Connectix*, a third-party reverse engineered Sony's PlayStation code not to release an unauthorized game, but instead to release an unauthorized emulator—a program allowing users to run PlayStation games on their computers without buying the console.[247] The Ninth Circuit once again upheld fair use, employing reasoning similar to *Sega*'s.[248] The decision's attempt to wedge the holding into the transformative-use paradigm was

---

240 *See infra* notes 251–54 and accompanying text.

241 Sega Enters. v. Accolade, Inc., 977 F.2d 1510 (9th Cir. 1992).

242 4 NIMMER & NIMMER, *supra* note 9, at § 13F.14[C][2].

243 *Sega*, 977 F.2d at 1514.

244 *Id.* at 1514–15. *But see id.* at 1523 (questioning the market impact of one game on another because "a consumer might easily purchase both").

245 *Id.* at 1527–28. The Federal Circuit in Atari Games Corp. v. Nintendo of Am. Inc., 975 F.2d 832 (1992), reached essentially the same conclusion in dicta one month prior. *See id.* at 843 ("Reverse engineering, untainted by the purloined copy of the 10NES [lockout] program and necessary to understand 10NES, is a fair use.").

246 4 NIMMER & NIMMER, *supra* note 9, at § 13F.05[C][1].

247 Sony Comput. Ent., Inc. v. Connectix Corp., 203 F.3d 596, 599 (9th Cir. 2000).

248 *See* 4 NIMMER & NIMMER, *supra* note 9, at § 13F.14[C][3].

unconvincing. It focused on the alleged transformativeness of the emulator due to its existence as a "wholly new product" that allows "opportunities for game play in new environments."[249] It did not articulate a differentiation of purpose because understandings of transformative use had not yet coalesced around that question.[250]

The entanglement approach would seem to cut against fair use for these reverse engineering cases insofar as the final products overlap in purpose with the works copied. In *Sega*, Accolade deployed the lock-out code to release games that worked with the Sega Genesis console, in line with Sega's own use of the code.[251] *Connectix* likewise involved a scenario in which the defendant used Sony's code to run PlayStation games, just as Sony had.[252] The entanglement approach would also be odd for these cases because the ultimate use was not the act being challenged. Reproduction of the copyright owner's expression occurred only during literal copying at the time reverse engineering took place, meaning it was the only act of copying that had to be adjudicated as fair or unfair. The products that the reverse engineers later released contained none of the plaintiffs' protected expression.

Disaggregation leads to a different puzzle because the copying had multiple purposes, some of which facially overlapped with the purpose of the original works. It is possible to reconstruct these cases to make them consistent with a purpose-oriented vision of transformative use. To do so, we would focus on the defendant's higher order objectives: Sega programmed its lock-out code to achieve a specific technical function, whereas Accolade copied Sega's code to achieve interoperability.[253] This purpose is plausibly distinctive. However, this reconstruction ignores the fact that achieving interoperability also requires duplicating the originally intended technical function.[254] The tension between the higher and immediate purposes of copying leaves disaggregation indeterminate.

---

249 *See Connectix*, 203 F.3d at 606–07.

250 As a 2000 decision, it preceded the 2006 sea change. *See supra* note 197 and accompanying text.

251 Sega Enters. v. Accolade, Inc., 977 F.2d 1510, 1515–16 (9th Cir. 1992).

252 *Connectix*, 203 F.3d at 599.

253 4 NIMMER & NIMMER, *supra* note 9, at § 13F.14[D][2].

254 As Justice Breyer observed in a later software case, "virtually any unauthorized use of a computer program . . . would do the same." Google LLC v. Oracle Am., Inc., 593 U.S. 1, 30 (2021).

How, then, do we reconcile these cases with the rest of fair use? One answer would be to carve them off as a separate category apart from transformative use. Factor two draws a line between traditional creative works and functional works,[255] so perhaps software's functional nature calls for different rules. But that move misses the nature of the works and the copying in the actual cases. Sega's games did not consist primarily of non-protectable material. Notwithstanding their inclusion of computer code,[256] the games were complex works consisting of myriad audiovisual and software elements.[257] Many of these elements were copyrightable even though segments like the lock-out code were not. Factor two was important not because games were functional *in toto*, but because the reverse engineer homed in on elements that were functional and therefore unprotectable. Traditional creative works like novels and paintings are not so different—they too contain a mix of protectable and unprotectable elements.[258] *Sega*'s logic may translate beyond software when a user copies unprotectable elements of these works.[259]

We arrive at a better answer if we consider what the transformative-purpose inquiry really does. One of its important roles is to screen for whether the user has intruded on a market that belongs to the copyright owner.[260] When the purpose is clearly distinct, the inquiry signals that the defendant's use sits outside the work's existing or customary markets.[261] But purpose is only a proxy for this substantive question. When purpose cannot provide decisive or satisfying answers, it is productive to consider market intrusion directly.

---

255 *See id.* at 22.

256 *Sega*, 977 F.2d 1510, 1526 (9th Cir. 1992) ("afford[ing] them a lower degree of protection than more traditional literary works").

257 A case decided the same year recognized "[t]he hallmark of a video game is the expression found in 'the entire effect of the game as it appears and sounds,' its 'sequence of images.'" *See* Atari Games Corp. v. Oman, 979 F.2d 242, 245 (D.C. Cir. 1992) (Ginsburg, J.) (internal citations omitted).

258 *See* BJ Ard, *Hybrid Innovation Regimes: The Interplay of IP and Non-IP Protections*, 109 IOWA L. REV. ONLINE 148, 150 (2024) ("Creative works in each field feature a distinctive combination of elements. As Seaman and Tran observe in gaming, some elements in other fields are covered by IP and others are not.") (citation omitted).

259 *Cf.* Sag, *supra* note 42, at 1904 (generalizing *Sega* in terms of non-expressive use).

260 *See* Andy Warhol Found. v. Goldsmith, 598 U.S. 508, 555 (2023) (Gorsuch, J., concurring).

261 *See supra* note 196 and accompanying text.

C. Substitutability

One could sidestep the difficulties of assessing purpose by placing greater emphasis on substitution and substitutability. Factor four already asks about the use's impact on the market for a copyrighted work,[262] and courts considered it the most important factor prior to the rise of transformative use.[263] More recently, the *Warhol* Court brought the same inquiry to bear on transformativeness under factor one, explaining that "the first factor considers whether and to what extent an original work and secondary use have substitutable purposes."[264] Although some dispute whether the two portraits were in fact substitutable,[265] the basic logic is sound: evidence showing that the use can substitute for the original undermines the claim that the use serves a distinct purpose.[266]

But what if competitive harm stems from the copying of unprotected elements? Prior cases have seldom had to confront whether that substitution is equally decisive. *Google Books*, as the most prominent case to address the matter, carries limited weight because the decision was almost too easy: the use was transformative and the court predicted few actual harms,[267] meaning little rode on its further exclusion of harms due to the copying of facts.[268] Generative AI may pose harder questions because it can eschew the copying of protected expression yet still pose risks of substitution. The closest guidance comes from a line of reverse-engineering cases that endorsed copying unprotected elements to create competing products. Determining how *Google Books* and these reverse-engineering cases transpose to generative AI is crucial to structuring a viable fair-use framework.

---

262 17 U.S.C. § 107(4).

263 4 NIMMER & NIMMER, *supra* note 9, at § 13F.08[G].

264 *Warhol*, 598 U.S. 508, 536 n.12 (2023). There is potential tension between this point and *Warhol*'s embrace of disaggregation, *see supra* note 223 and accompanying text, because focusing on substitutability suggests a holistic approach focused on the ultimate market-facing use.

265 *See* 598 U.S. at 567 (Kagan, J., dissenting) ("You would see them not as 'substitute[s],' but as divergent ways to (in the majority's mantra) 'illustrate a magazine about Prince with a portrait of Prince.' Or else you (like the majority) would not have much of a future in magazine publishing.") (internal citations omitted).

266 4 NIMMER & NIMMER, *supra* note 9, at § 13F.11[E].

267 *See supra* notes 208–09 and accompanying text.

268 *See infra* notes 280–82 and accompanying text.

1. *Weighing Different Harms*

Warhol's unvarnished statement that substitutability undermines transformativeness[269] raises difficulties.[270] If all substitution counts against transformativeness, then it follows that generative art systems are not transformative when they generate works that can substitute for the works on which they are trained.[271] Yet copyright ostensibly does not protect non-expressive or low-expression elements like facts, ideas, or tropes.[272] *Warhol*'s formulation of the substitutability test would cut against fair use for generative AI systems whose outputs displaced human artists even if the systems studiously avoided replicating the protected, expressive elements of any training work.

We could rehabilitate *Warhol*'s statement of the test by importing a limiting principle already developed in factor four jurisprudence.[273] Factor four excludes market harm due to copying of unprotected elements, and it stands to reason that the same limitation should carry over to substitutability under factor one.[274] To be sure, courts rarely have occasion to articulate this point because an infringement claim will fail without recourse to fair use if the defendant copies only unprotected elements.[275] Cases going back to the foundational *Baker v. Selden* decision[276] have denied infringement claims directed at competitors who used unprotected material like accounting methods,[277]

---

[269] 598 U.S. 508, 536 n.12 (2023).

[270] 4 Nimmer & Nimmer, *supra* note 9, at § 13F.10[G][2][c][iii]. For further scrutiny of the Court's formulation, see Glynn S. Lunney, Jr., *Transforming Fair Use*, 14 N.Y.U. J. I.P. & Ent. L. 169, 197 (2025) ("[T]hat conclusion is nonsense, and the reasoning by which the Court reached it misses the point entirely.").

[271] *See infra* subpart IV.B.

[272] *See* Jeanne C. Fromer, *An Information Theory of Copyright Law*, 64 Emory L.J. 71, 97–99 (2014) (explaining the idea–expression distinction alongside a "collection of doctrines" that "liberates precisely those aspects—ideas, facts, stock elements—that might readily be buried under the noise of expression in any one particular work"); Van Houweling, *supra* note 31, at 106–07 (developing a taxonomy of exclusions for ideas, methods, and facts).

[273] *See* 4 Nimmer & Nimmer, *supra* note 9, at § 13F.08[C][2].

[274] *Id.*

[275] *Id.*

[276] 101 U.S. 99 (1879).

[277] *Id.* at 104 (holding that copyright in a book on accounting could not prevent others from publishing books using the same accounting method).

recipes,[278] or yoga sequences[279] to the their creators' detriment. There was no need to consider fair use as a defense in the absence of a colorable infringement claim, and thus no occasion to weigh the resulting market harms.

Google Books provides the clearest expression of the exclusion, albeit as a fallback position after the court downplayed the likelihood of market harm.[280] For the most part, Google Books did not displace book purchases. Although it provided book snippets, the snippets it provided per book were few and non-contiguous.[281] This meant most people who wished to read the book had to obtain copies through standard channels. Many technological uses previously upheld as transformative and fair possessed similar safeguards against public use.[282] The question of which substitutions should count arose only in a more specific context: *Google Books* recognized that snippets might substitute for book sales if someone were seeking a specific uncopyrightable fact, like the year President Roosevelt contracted polio.[283] The court declined to hold this market harm against Google because it involved only substitution as to an unprotected fact.[284] To count against fair use, the court held that substitution had to implicate an author's protected expression.[285]

2. *Legitimate Competition*

Stronger precedent comes from harder cases where courts excused the copying of unprotected elements even when it

278 *See* Publ'ns Int'l, Ltd. v. Meredith Corp., 88 F.3d 473, 482 (7th Cir. 1996) (holding analogously with respect to a cookbook and a competitor's use of the same recipes).

279 Bikram's Yoga College of India, L.P. v. Evolation Yoga LLC, 803 F.3d 1032, 1038 (9th Cir. 2015) (holding analogously with respect to a yoga book and a competing studio's use of the same sequence of poses).

280 *See* Authors Guild v. Google Inc. (*Google Books*), 804 F.3d 202, 224 (2d Cir. 2015).

281 *See id.* at 209–10.

282 For example, prior image search engines harvested images from the internet and copied them in full but then discarded the images after creating low-resolution thumbnails. *See* Perfect 10, Inc. v. Amazon.com, Inc., 508 F.3d 1146, 1165 (9th Cir. 2007); Kelly v. Arriba Soft Corp., 336 F.3d 811, 819 (9th Cir. 2003). Although the search engines made the thumbnails publicly available, the thumbnails' low resolution rendered them largely inadequate to serve the aesthetic purposes of the originals. *See* 336 F.3d at 819. These design decisions mitigated the risk of the thumbnails inflicting substitutionary harm.

283 804 F.3d at 224.

284 *Id.*

285 *Id.*

led to plausible market harm. Here we return to the paradigmatic reverse-engineering case *Sega*.[286] Reverse engineering may struggle to sustain transformation of purpose because software copying typically aims, at some level, to duplicate the original work's function.[287] Moreover, the potential competitive harm of defendant's copying was real: it fueled the creation of new games that competed with Sega's licensed offerings.[288] The defendant nonetheless prevailed because its duplication of unprotected functional elements did not implicate the sorts of harms that copyright cares about.[289] Per the court, the defendant "sought only to become a legitimate competitor" in the relevant market and had not copied the protected elements that determine a game's commercial success.[290]

The Supreme Court reaffirmed *Sega*'s approach in its 2021 *Google v. Oracle* decision.[291] In doing so, it established that alleged losses in the billions of dollars can be discounted for copying that focuses on functional elements.[292] Admittedly, the opinion's position on the exclusion of unprotected elements is muddled. The alleged infringement was Google's copying of over 10,000 lines of Sun Java's declaring code to create a coding platform for its Android smartphone operating system.[293] The Court's determination that the code was closer to unprotected than most software[294] colored the rest of the opinion,[295] but the Court did not expressly bring the point to bear on factor four. It instead side-stepped the issue of market harm by adopting the disputable premise that Google had not in fact

286 Sega Enters. v. Accolade, Inc., 977 F.2d 1510 (9th Cir. 1992); *see supra* section II.B.3. One complication of reverse-engineering, in contrast to cases dealing with copying the method of an accounting book or cookbook, is that it is impossible to extract the unprotected elements without also copying the expression at an intermediate phase. *See infra* notes 308-09 and accompanying text.

287 *See supra* note 254.

288 977 F.2d at 1523. The court discounted the likely competitive harm, however, on the rationale that "video game users typically purchase more than one game." *See id.*

289 *See* Sag, *New Legal Landscape*, *supra* note 31, at 311.

290 977 F.2d at 1523.

291 Google LLC v. Oracle Am., Inc., 593 U.S. 1 (2021).

292 *Id.* at 43 (Thomas, J., dissenting). I cite Justice Thomas's dissent for its repeated insistence that this "unlikely result" was remarkable given the amount copied and the apparent market impact. *See id.* at 43.

293 *See id.* at 30–34 (majority).

294 *Id.* at 28–29.

295 *Id.* at 52 (Thomas, J., dissenting) ("This opening mistake taints the Court's entire analysis.").

harmed the licensing market for the plaintiff's platform.[296] It cited *Sega* and *Connectix* with approval,[297] but for the point that public benefits—presumably, the benefits of legitimate competition—bear on factor four.[298]

Google nonetheless provides striking precedent for upholding fair use despite plausible substitutionary harms. And it does so in a way that clashes with the transformative-purpose inquiry: Google copied Sun Java to invoke the same functions for which it was coded and to create a similar platform. The coherent explanation is that Google's use was fair because it did not compete in any market that rightly belonged to the plaintiff. This brings us to the further question of determining which markets belong to the plaintiff and which do not. For guidance, we turn to theories grounded on the divide between expressive and non-expressive use[299] before developing a complementary perspective on the exploitation of authorial versus non-authorial value.[300]

D. Non-Expressive Fair Use

Non-expressive fair use may yield distinct insights for generative AI. Historically, non-expressive use has provided alternative, complementary grounds for explaining decisions articulated in terms of transformative use.[301] Broadly speaking, the argument is that uses are fair so long as they do not exploit a work's protected expression.[302] Non-expressive use may have unique importance for analyzing generative AI to the extent that conventional approaches, like transformativeness, ask the wrong questions.[303] Proponents and critics have accordingly begun the work of extending and testing the framework in this context.[304]

---

[296] *Id.* at 35–36 (majority); *see id.* at 52–56 (Thomas, J., dissenting) (decrying these findings as implausible); *see also* 4 NIMMER & NIMMER, *supra* note 9, at § 13F.03[C][5][d] (unpacking the Court's treatment of implicit jury findings).

[297] 593 U.S. at 22, 39 (majority).

[298] *See* 4 NIMMER & NIMMER, *supra* note 9, at § 13F.08[D][2].

[299] *See infra* subpart II.D.

[300] *See infra* Part III.

[301] *See* Sag, *supra* note 11, at 1675 ("[T]he transformative use doctrine is but one manifestation of the broader principle of expressive substitution.").

[302] *Id.* at 1608–09; *see* Brauneis, *supra* note 9, at 22–28 (comparing theories).

[303] *See supra* section II.B.3.

[304] *See, e.g.*, Sag, *supra* note 9, at 309 ("For the most part, the copyright implications of the new wave of LLMs are no different from earlier applications of text data mining."); Jacob Alhadeff, Cooper Cuene & Max Del Real, *Limits of*

But meeting this challenge requires confronting unanswered questions. Foremost among them is the matter of what it means for copying to be non-expressive.[305] Paradigmatic examples of non-expressive use like the Google Books service undeniably copied expressive content but refrained from circulating that expression;[306] they were also indifferent to the works' expression and did not result in the creation of new expressive works, much less competing works. Because Google Books was non-expressive along each of these dimensions, the theory could tolerate ambivalence regarding precisely which dimensions mattered. Generative AI is different: it forces us to choose which features matter because many instantiations of generative AI will fail one or more of these tests.

1. *Theories of Non-Expressive Use*

Theories of non-expressive use converge on the principle that a use is fair so long as it does not exploit a work's protectable expression.[307] Uses in this vein have often copied works in full and then processed or stored them to some further end. The principle has become salient because of how computers work. There would be no need for fair use if one hand-copied the facts of a book because facts, as such, are unprotectable.[308] Fair use becomes necessary to justify analogous computer uses, however, because digital utilizations typically must copy a work in full to extract anything even if the computer later deletes all expressive portions.[309]

Matthew Sag first introduced the theory in 2009 to bring coherence to early search-engine and reverse-engineering cases[310] and to provide guidance for the Google Books saga then unfolding.[311] Building on the precept that copyright aims

---

Algorithmic Fair Use, 19 WASH. J.L. TECH. & ARTS 1, 40–47 (2024) (contrasting generative AI with prior non-expressive uses); Opderbeck, *supra* note 9, at 976 (questioning the doctrinal basis for non-expressive use). Some also anticipated the discussion prior to the explosion of generative AI. *See generally* Lemley & Casey, *supra* note 31; Sag, *supra* note 11; Sobel, *supra* note 9.

305 Brauneis, *supra* note 9, at 3–4.

306 *See generally* Authors Guild v. Google Inc. (*Google Books*), 804 F.3d 202 (2d Cir. 2015); *see also supra* sections II.B.1 & II.C.2 (assessing the service's transformativeness and market impact).

307 *See supra* note 302 and accompanying text.

308 *See* Sag, *supra* note 11, at 1613.

309 *Id.*

310 *See id.* at 1618–24, 1654–55.

311 *See id.* at 1643–44.

specifically to protect against expressive substitution, Sag concluded that "acts of copying that do not communicate the author's original expression to the public" are fair.[312] Later text-and-data-mining projects provided strong examples: the projects might copy and store full texts, but their statistical outputs could not substitute for the original texts' expression because they conveyed no expression to the user.[313]

Sag's focus on expressive substitution positioned him to consider which markets copyright recognizes. The answer surely could not be that all lost sales count: the Supreme Court had already recognized that copyright owners cannot complain of harm that flows from criticism or licenses foregone in the market for parodies.[314] Cases like *Sega* suggested that lost sales flowing from use of non-expressive elements to create new, competing expression should not count either.[315] Because the uses under study produced practically no market harm, however, the cases did not demand close scrutiny; Sag's work moved from the market-identification problem to the stronger conclusion that "non-expressive uses *categorically* do not" pose a threat of expressive substitution.[316]

The non-expressive use theory gained steam from *Google Books*' recognition of fair use in mass commercial copying to create a searchable index,[317] leading to further elaboration.[318] Some scholars, like James Grimmelmann, cautioned against the excesses of presumptively treating robot reading as fair.[319] Others pushed the theory further. Applying the principle to generative AI, Oren Bracha argues that copying that does not convey a work's expression to a human observer should not count as infringement at all.[320]

Mark Lemley and Bryan Casey articulate an alternate approach—"fair learning"—that might be characterized as

---

312 *Id.* at 1625.

313 Sag, *supra* note 31, at 327–28.

314 Sag, *supra* note 11, at 1654 (discussing Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 591–92 (1994)).

315 *See id.* at 1655; *see also supra* section II.C.2 (discussing Sega Enters. v. Accolade, Inc., 977 F.2d 1510 (9th Cir. 1992)).

316 Sag, *supra* note 31, at 327.

317 *See generally* Authors Guild v. Google, Inc. (*Google Books*), 804 F.3d 202 (2d Cir. 2015).

318 *See* Brauneis, *supra* note 9, at 3 n.1 (assembling the literature).

319 *See* James Grimmelmann, *Copyright for Literate Robots*, 101 Iowa L. Rev. 657 (2016).

320 *See* Bracha, *supra* note 9, at 181.

"non-use of expression."[321] They focus not on whether the use communicates expression to the public, but instead on which aspects of the work the defendant utilizes.[322] Copying to appropriate the plaintiff's expression would cut against fair use, whereas copying to obtain unprotected facts and ideas would favor it. Lemley and Casey give the example of feeding pictures of stop signs to a computer so that it can learn how stop signs look in a variety of settings and at different angles.[323] Going further than other versions of non-expressive use, fair learning would also endorse public-facing uses where the audience is interested in facts as opposed to "the copyrighted bits" like "accidents of the plaintiff's angles and lighting."[324] This underscores their agreement with cases following this rationale, like one upholding circulation of the Kennedy assassination film as fair,[325] and their disagreement with the larger trend toward rejecting fair use for the redistribution of newsworthy photographs and videos[326] or the copying and circulation of scholarly journals.[327]

These theories have largely overlapped with transformative use in practice. Differentiation in purpose was enough to support fair use for a range of prior copy-reliant technologies,[328] barring reverse engineering,[329] without having to dig into whether they exploited the works' expression. Strong justifications for allowing market harm were not required because actual market harm was minimal.[330] Generative AI now poses harder questions: non-expressive use may take on unique importance to the extent transformative purpose is indeterminate and market harm is plausible.

---

321 *See generally* Lemley & Casey, *supra* note 31.

322 In doing so, they broach the value question addressed below: "It's fair because the value the ML system gets from the copyrighted work stems from the part of the work the copyright law has decided belongs to the public, not to the copyright owner." *Id.* at 779.

323 *Id.* at 749 & n.34.

324 *Id.* at 781.

325 *See id.* at 781–82, 782 & n.207 (discussing Time Inc. v. Bernard Geis Assocs., 293 F. Supp. 130 (S.D.N.Y. 1968)).

326 *See id.* at 782.

327 *See id.* at 780 (discussing Am. Geophysical Union v. Texaco, Inc., 60 F.3d 913 (2d Cir. 1994)).

328 *See supra* sections II.B.1–2.

329 *See supra* section II.B.3.

330 *See supra* subpart II.C.

2. *Defining Non-Expressive Use*

Non-expressive use faces definitional challenges as commentators seek to extend it to generative AI. Robert Brauneis crystallizes the point in work questioning whether the theory could justify training if non-expressive use were defined with greater precision.[331] The taxonomy he develops illuminates three possible meanings. One might focus on *constitutive expression* and define non-expressive use as any use that seeks to extract facts or ideas from a work while remaining indifferent to the expression that constitutes the work.[332] For example, one might deploy machine learning on street photographs to identify what stop signs look like while designing the process to ignore or discard the photographs' creative features.[333] Alternatively, one might focus on *actionable expression*, defining non-expressive use as any that falls short of publicly reproducing the work.[334] This approach would embrace use of copyrighted works as inputs to machine learning so long as the resulting system did not reconvey the works' expression to later users in a manner that itself infringed.[335] Finally, one might focus on *felt expression*, defining non-expressive use as any use that is not emotionally experienced by a human.[336]

Precision was not required in prior cases because they often satisfied all three tests. Take the anti-plagiarism tool at issue in *iParadigms*.[337] Although the defendant's TurnItIn system copied student essays to a private server for comparison with later submissions,[338] TurnItIn was indifferent to each essay's actual content, meaning it did not use constitutive expression. The system neither distributed the originals nor created new ones, meaning it did not yield actionable expression. Moreover, the absence of any human readers precluded felt expression.

Generative AI now forces the question because aspects of systems and their training may fail one or all the tests. Training may plausibly seek to extract or learn to reconstitute the expressive elements that embody something like Picasso's

---

331 *See* Brauneis, *supra* note 9, at 4.

332 *Id.* at 22–26.

333 *Id.* at 26.

334 *Id.* at 26–27.

335 *See id.* at 27.

336 *Id.* at 27–28.

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

338 *Id.*

style, implicating constitutive expression.[339] Any reconstitution of expressive elements would be public facing, leading to actionable expression if it crossed the threshold of substantial similarity.[340] And human audiences are meant to enjoy the expressive content of these outputs, which may implicate the training works' felt expression. The fit of the non-expressive label may vary depending on which definitions we use and the particulars of the systems in question.

3. *Returning to Market Impact*

Non-expressive use theories also stand in a complex relationship with market substitution. On one hand, the theories treat substitution as a central concern and seek to explain why non-expressive uses do not implicate categories of market harm that matter for copyright. Sag limits relevant harms to expressive substitution and posits that non-expressive uses "categorically do not" lead to this sort of substitution.[341] In a separate analysis, Sag also frames factor three's analysis of the amount and substantiality of copying around whether the use "substitute[s] for the expressive value of the author's original expression," concluding that non-expressive uses generally do not.[342] Lemley and Casey likewise define unfair uses as those that exploit "the creative aspects of the work that copyright values"[343] and discounts any harms from copying "the part of the work the copyright law has decided belongs to the public."[344] These formulations imply a strong shield against legal liability notwithstanding factual harm.

On the other hand, however, much writing applying non-expressive use to generative AI urges the use of balancing tests that weigh competitive harms rather than fully embracing the implied liability shield.[345] Sag's latest work opens the analysis

339 *See supra* section I.D.1.

340 *See supra* section I.D.2.

341 Sag, *supra* note 31, at 327 (emphasis omitted).

342 Sag, *supra* note 11, at 1652–53.

343 Lemley & Casey, *supra* note 31, at 777.

344 *Id.* at 779.

345 This move aligns non-expressive use scholarship with pragmatic approaches proposed by commentators coming at the problem from other angles. *See, e.g.*, Ian Ayres & Jack M. Balkin, *The Law of AI Is the Law of Risky Agents Without Intentions*, U. CHI. L. REV. ONLINE, 2024, at *1, *10 ("[T]he law should require, as a condition of a fair use defense, that AI companies take a series of reasonable steps that reduce the risk of copyright infringement, even if they cannot completely eliminate it.").

to consider "whether the challenged use undermines the economic incentives that copyright is designed to create, even in the absence of direct expressive substitution."[346] Lemley and Casey previously conceded that "fair use is unlikely to save" certain uses that copy the style of established artists—partly for concern with risks of expressive substitution[347] and partly in recognition that AI uses will receive little sympathy in the courts.[348]

This Article reframes the balancing approach through the lens of authorial value. While non-expressive use theories have typically emphasized either factor one's purpose-and-character test or factor two and three's inquiry into the nature and extent of the elements copied, this analysis pivots to factor four. It extends fair use to copying that primarily exploits a work's non-authorial value. This shift necessitates moving beyond questions of whether the work is recirculated and which parts are used to also consider the use's context and market impacts. Although this perspective generally supports the fair use of non-expressive elements, the two approaches do not overlap perfectly. As the next Part demonstrates, analysis of non-authorial value extends more readily to pragmatic balancing.

III
Non-Authorial Value and Fair Use

Copyright has a latent space all its own. Courts and scholars do not literally encode cases to interpolate the correct principles for copyright adjudication. However, doctrinal analysis is the process of mapping the relationship between cases to identify principles that make them coherent. As with generative art, the sorts of answers we obtain depend largely on our prompts—the questions we pose. The principles we decode when *Warhol* pushes us to inquire about the parameters of substitutability may accordingly lead us to frame our answers differently and to see trends otherwise obscured.

Examining the courts' treatment of substitution effects reveals an under-appreciated distinction in caselaw based on

346 Sag, *supra* note 42, at 1917. Grimmelmann, who was already skeptical of extensive fair use for robots, likewise urges nuanced analysis in recent co-authored work. *See* Lee, Cooper, & Grimmelmann, *supra* note 23, at 110 ("This categorical argument does not work for generative-AI models that can generate expressive works.").

347 Lemley & Casey, *supra* note 31, at 778 (explaining that "some purposes . . . seem more substitutive than transformative").

348 *See id.* at 746.

whether a use exploits a work's non-authorial value. Copyright protects copyright owners against intrusion upon authorial value—the portion of a work's value that flows from the original expression contributed by its author. However, the following discussion shows that copyright law has historically allowed others to exploit a work's non-authorial value—that which flows from the work's non-original or non-expressive elements, its use of tropes that derive value from societal expectations rather than the author's creative choices, and in some instances from third-party contributions. This treatment sometimes comes through in fair use and sometimes through doctrines that simply exclude the exploitation of non-authorial value from liability. Training for generative AI may find shelter under this principle depending on precisely which types of value it exploits.

The key distinction between the non-authorial value account and prior non-expressive use theories is its focus on addressing substitutability under factor four (and *Warhol*'s approach to factor one).[349] Prior theories addressed other salient questions and remain important to the analysis of other factors. Characterizing the use as non-expressive in the context of factor one was responsive to the need for concepts other than transformativeness to describe the purpose and character of a valid use, especially when the transformative label did not fit.[350] Likewise, focusing on whether that which was used was expressive in nature, in the context of factors two and three, was responsive to the ongoing difficulty of implementing copyright's limiting doctrines for digital technologies that copied works in full as a prerequisite to further use.[351] The distinction between authorial and non-authorial value is responsive to the problem of working through these limiting doctrines in the substitutability framework articulated by *Warhol*.

---

349 *See supra* subpart II.C.

350 *See, e.g.*, Sag, *supra* note 11, at 1647 ("It would be better to recognize that uses which do not relate to the expressive appeal of a work may find favor under the first fair use factor—whether they qualify as transformative in the expressive sense or not.").

351 The Supreme Court's decision in *Google LLC v. Oracle America, Inc.*, 593 U.S. 1 (2021), provides fresh support for this position by analyzing factor two specifically in context of the elements used. 4 NIMMER & NIMMER, *supra* note 9, at § 13F.06[A][2] ("[W]hat mattered for factor two was not the nature and protectability of the work as a whole, but rather the nature and protectability of those portions copied.").

One could approach substitutability by focusing on whether a use exploits *expressive* value.[352] The further move to focus on *authorial* value aims to clarify and broaden the inquiry.[353] It also takes inspiration from the parallel discussion of authorship in AI copyrightability debates. The problem with the registration of AI outputs has not been that they lack expressive content. Instead, the argument against copyrightability is that AI users lack authorship for failure to conceive of or control the execution of that expression.[354] Echoing *Warhol*'s focus on uses rather than works,[355] this line of analysis pushes us to consider context beyond that which is evident from dissecting a work on its face.

This broader lens helps better explain the scope and the limitations of the freedom to exploit non-authorial value outlined below. The idea–expression dichotomy can explain many cases and doctrines allowing for the exploitation of non-authorial value because, as a matter of copyright, the author can claim only the value of the work's expression.[356] But it bears note that copyright does not secure the entire value of a

---

352 *See supra* notes 341–44 and accompanying text.

353 "Expressive value" is also slippery as a term. Although it could refer to the market value attributable to a work's expression, it could also refer more intangibly to a work's aesthetic quality. We adopt this usage when we say, for example, that a computer program or a factual biography has little expressive value relative to an oil painting. I seek to move away from that ambiguity and to re-center the discussion on the economic value that rightly does or does not belong to the author.

"Authorial" terminology nominally refers to the conception and control-over-execution requirements of copyright authorship, *see infra* note 354 and accompanying text, but has the added advantage of subsuming copyright's doctrinal limits around originality and expression because copyright authorship is deeply entwined with these concepts. As Dan Burk put it, "Copyright authorship requires both an act—the act of fixing expression in a tangible medium—as well as a type of mental effort or creative activity to originate the expression that is fixed." Dan L. Burk, *Thirty-Six Views of Copyright Authorship, by Jackson Pollock*, 58 HOUS. L. REV. 263, 268 (2020).

354 *See* Thaler v. Perlmutter, 687 F. Supp. 3d 140, 146 (D.D.C. Aug. 18, 2023) (appeal filed Oct. 18, 2023); 5 NIMMER & NIMMER, *supra* note 9, at § 20.05[B][1]. The apparent existence of works within a system's latent space prior to any prompting also raises questions for originality because the outputs may not originate with the user. *Id.* at § 20.05[B][3]. *See also supra* section I.D.2 (explaining how outputs may be predetermined as of the time of training).

355 *See supra* notes 223–24 and accompanying text.

356 *See id.* § 102(b); Golan v. Holder, 565 U.S. 302, 328–29 (2012). Use of "idea–expression" terminology serves here as shorthand for a broader protectable versus unprotectable distinction that also encompasses an expression–fact and expression–functional-element divide. *See* Pamela Samuelson, *Why Copyright Law Excludes Systems and Processes from the Scope of Its Protection*, 85 TEX. L. REV. 1921, 1923 & n.11 (2007)

work's expression to its author.[357] At most, it secures the value of whatever portion of the work's expression stems from its author's original contributions.[358] Authorship and originality give us language to articulate this distinction in scenarios where the divide between idea and expression alone may not capture it. This is not to say that the focus on non-authorial value uniformly expands fair use. Moving away from binary determinations around expression instead forces us to grapple substantively with hard cases where exploitation of non-authorial value sits close to destruction of the work's entire market value.

A. Third-Party Contributions: *Google*'s Reimplementation of Fair Use

Google v. Oracle[359] provides a striking endorsement for the freedom to exploit non-authorial value contributed by third-party investment and related network effects notwithstanding the apparent risk of market harm. The *Google* Court upheld as fair Google's copying of over ten-thousand lines of declaring code to reimplement the Java API on the Android smartphone platform.[360] Much commentary on the case has focused on its treatment of transformativeness (tortured)[361] or the significance of the work's functionality under factor two (pervasive).[362] Here I highlight another cross-cutting point: the determination that Sun Java's declaring code derived value largely from third-party investment.

The Court carefully separated whatever value was inherent to the code from that which was contributed by others. In

---

357 Nor should it. IP scholarship identifies spillovers—the positive externalities that IP creators cannot internalize—as a desirable feature of our IP regimes. For the leading account, see Brett M. Frischmann & Mark A. Lemley, *Spillovers*, 107 COLUM. L. REV. 257 (2007).

358 Putative authors must exclude elements that are not their own original authorship at the copyright registration stage, *see Copyright Registration Guidance: Works Containing Material Generated by AI*, 88 Fed. Reg. 16190, 16192 (Mar. 16, 2023) (stating the Copyright Office's position), and any such elements are also filtered out during infringement analysis, *see* Christopher Jon Sprigman & Samantha Fink Hedrick, *The Filtration Problem in Copyright's "Substantial Similarity" Infringement Test*, 23 LEWIS & CLARK L. REV. 571, 572–73 (2019).

359 593 U.S. 1 (2021).

360 *See id.* at 29–35.

361 The dissent voices the concern: "Now, we are told, 'transformative' simply means—at least for computer code—a use that will help others 'create new products.' That new definition eviscerates copyright." *Id.* at 58 (Thomas, J., dissenting) (internal citation omitted).

362 In the same spirit, the dissent concludes that the majority's treatment of factor two "taints the Court's entire analysis." *Id.* at 52.

assessing the nature of the work under factor two, it observed: "Unlike many other programs, [Sun Java's] value in significant part derives from the value that those who do not hold copyrights, namely, computer programmers, invest of their own time and effort to learn the API's system."[363] In the context of market effects under factor four, it noted "[t]his source of Android's profitability has much to do with third parties' (say, programmers') investment in Sun Java programs. It has correspondingly less to do with Sun's investment in creating the Sun Java API."[364] The fact that Google sought to make use of value contributed by third parties' investments in the platform, not the plaintiff's authorial contributions, thereby favored fair use under both factors.

On the surface, the decision seems to penalize the copyright owner for the work's popularity. But the deeper rationale is its refusal to allow a party to use copyright to monopolize contributions made by third parties, including the extensive network effects flowing from widespread adoption. Google's platform certainly outcompeted Oracle's and may have cost it licensing fees.[365] The Court nonetheless discounted the harm because the value Oracle lost stemmed from others' investments.

The logic is even more straightforward for reverse-engineering cases like *Sega*.[366] The fact that Accolade sought and retained only the uncopyrightable lock-out code[367] was a strong indicator that it did not seek to exploit the authorial value bound up with a work's expression. Context shows that Accolade did not seek to exploit the value of Sega's game at all: it wished to tap into the market for a popular video game console by releasing games of its own.[368] Like Google many years later, Accolade sought to exploit network effects attributable to third-party interest. Whereas *Sega* is often discussed in the context of technical interoperability,[369] *Google* signals an expansion of *Sega*'s logic beyond technical interoperability

---

363 *Id.* at 28–29 (majority).

364 *Id.* at 39.

365 *Id.* at 53 (Thomas, J., dissenting).

366 Sega Enters. Ltd. v. Accolade, Inc., 977 F.2d 1510 (9th Cir. 1992). *See supra* section II.B.3.

367 *See* 977 F.2d at 1521–22.

368 *Id.* at 1523.

369 *See, e.g.*, Mark A. Lemley & Pamela Samuelson, *Interfaces and Interoperability After* Google v. Oracle, 100 TEX. L. REV. 1, 25–26 (2021) (noting "strong consensus . . . that copyright did not give programmers the ability to prevent others from reusing parts of another's program when necessary to enable compatibility" but "less agreement on the legal doctrine that compelled this result").

to a greater range of scenarios where copyright owners might attempt to monopolize value not attributable to their own creative investments.

Copyright's non-protection of third-party contributions is not limited to fair use. Take copyright's refusal to protect games and game rules.[370] Bruce Boyden has theorized that this exclusion stems from recognition that games are tools through which game players contribute expression rather than vehicles for the game makers' own expression.[371] This refusal accordingly makes sense from the perspective of denying control over non-authorial value: much of the game's value derives from third parties rather than the game maker's authorship.

B. Exploitation of Facts and Unprotected Elements

Copyright also vindicates the exploitation of non-authorial value by recognizing freedoms to exploit facts and other unprotected elements. The literature is replete with accounts of how free use of these elements enriches the public domain and advances copyright's expressive goals through operation of the idea–expression distinction and related doctrines like merger.[372] This facet of copyright law has remained latent in fair use, however, because many of the cases where it applies do not reach fair-use determinations.[373] Sometimes, copying of unprotected elements falls short of interfering with authorial value because the elements literally do not originate with the author. No matter how painstaking the creative process, the author does not create the facts a work conveys. They preexist the authorial process. Reproducing conventional facts, like dates, may lead to market substitution, but the resulting harm is non-actionable because it diverts only the work's non-authorial value.

We see courts wrestle with this line in fair use as applied to less conventional facts. Cases like *Warhol*,[374] where the photographer has prevailed, have sustained criticism for ostensibly

---

370 Bruce E. Boyden, *Games and Other Uncopyrightable Systems*, 18 Geo. Mason L. Rev. 439, 440 (2011).

371 *Id.* at 442 ("Games . . . do not communicate expression to the players so much as provide a forum for the gameplay experience to occur.").

372 *See, e.g.*, Van Houweling, *supra* note 31, at 107 ("This is not because unprotected elements are not valuable enough to justify copyright, but rather because they are so valuable that they belong in the public domain.").

373 *Cf. supra* section II.C.1 (addressing the issue in the context of market substitution).

374 Andy Warhol Found. v. Goldsmith, 598 U.S. 508 (2023).

awarding protection over the look of a person's face. A prior suit in which a photographer sued over the use of President Obama's likeness in Shepherd Fairey's famous "Hope" poster provoked the same reaction.[375] Copyright simply does not endow photographers with the exclusive right to depict a person's face.[376]

Courts that have deemed the copying of photographic portraits unlawful have deflected the criticism by emphasizing that the infringement stemmed from exploitation of the authorial aspects of the photographs. The pivotal *Sarony* case establishing the copyrightability of photographs emphasized the photographer's choices with respect to artfully capturing the scene through "arranging and disposing the light and shade," or through composing the scene itself, by posing the subject, "selecting and arranging the costumes, draperies, and other various accessories," and "suggesting and evoking the desired expression."[377] The influential *Mannion* decision referred to these aspects as originality in "rendition" and "creation of the subject," respectively.[378] These are the sorts of contributions that add authorial value to a work. And, in the Second Circuit's view, these were the aspects that Warhol sought to exploit from Goldsmith's work: "Warhol's modifications serve chiefly to magnify some elements of [the photograph]," "down to the glint in Prince's eyes where the umbrellas in Goldsmith's studio reflected off his pupils."[379]

Cases going the other way have emphasized the opposite. Consider *Kienitz*, in which the defendant modified a mayoral photograph to print on a t-shirt.[380] Following heavy modification, including a flattening of depth and recoloring of the mayor's face to lime green, "the expression in [the mayor]'s eyes [could] no longer be read" and "the effect of the lighting in the original

---

375 *See* Justin Hughes, *The Photographer's Copyright — Photograph As Art, Photograph As Database*, 25 HARV. J.L. & TECH. 339, 390 (2012) ("In this particular case, Fairey clearly created a poster *based* on the photograph, but he did not copy any original elements of the photograph."). The dispute settled prior to any judicial determination. *See* Fairey v. Associated Press, No. 09-civ-01123 (S.D.N.Y. Feb. 9, 2009).

376 Hughes, *supra* note 375, at 377.

377 Burrow-Giles Lithographic Co. v. Sarony, 111 U.S. 53, 60 (1884).

378 Mannion v. Coors Brewing Co., 377 F. Supp. 2d 444, 452–54, (S.D.N.Y. 2005). *Mannion* also identified the potential for thin originality in timing. *Id.* at 452–53.

379 Andy Warhol Found. v. Goldsmith, 11 F.4th 26, 43, 48 (2d Cir. 2021), *aff'd*, 598 U.S. 508 (2023).

380 *See* Kienitz v. Sconnie Nation LLC, 766 F.3d 756, 756 (7th Cir. 2014).

[was] almost extinguished."[381] The defendant's eschewal of the photographer's creative contributions supported its fair-use victory.[382] Or consider the instructively bizarre example of *Blanch v. Koons*.[383] Jeff Koons began with a fashion photograph featuring a close-up of feet posed at a jaunty angle across a man's lap in what appears to be a first-class airplane cabin.[384] Koons took the feet entirely out of that setting (dangling them over piles of food), placed them at a different angle (simply pointing down), and altered their coloration.[385] Any value contributed by the plaintiff's posing of the subject and composure of the scene, or even her intended color scheme, was all but eliminated.[386]

C. Popular Expectations and Interest

Use of non-expressive elements generally does not intrude upon authorial value. To go one step further, the use of expression does not uniformly exploit authorial value. Sometimes the value of expression arises not from the author's original contributions, but from satisfaction of pre-existing stylistic conventions. Echoing the previous discussion of *Google v. Oracle*, the value of a work or some elements of a work may derive from third-party interest or investments. Prior decisions upholding the use of this sort of value as non-infringing or fair take on new coherence within the authorial value paradigm.

Scenes-a-faire doctrine is emblematic. Copyright does not protect the standard tropes and conventions audiences expect within a particular genre.[387] Inclusion of these elements is no bar to copyrightability—they still count as expressive—but infringement analysis often filters them out so they cannot establish actionable similarity.[388] Granted, many explanations for scenes-a-faire center on lack of originality rather than

---

381 *Id.* at 757, 759.

382 "Defendants removed so much of the original that, as with the Cheshire Cat, only the smile remains." *Id.* at 759.

383 467 F.3d 244 (2d Cir. 2006).

384 *Id.* at 248.

385 *Id.*

386 "As Blanch testified in her deposition, her key creative decisions in the shoot were the choice of an airplane cabin as a setting and her placement of the female model's legs on the male model's lap. But neither the airplane background nor the man's lap appear in 'Niagara.'" *Id.* at 258.

387 *See* Incredible Techs., Inc. v. Virtual Techs., Inc., 400 F.3d 1007, 1015 (7th Cir. 2005) (extending the doctrine to conventions of video-game medium and to depiction of real-world activity of golf).

388 Apple Computs., Inc. v. Microsoft Corp., 35 F.3d 1435, 1444 (9th Cir. 1994).

third-party expectations.[389] The point works well enough. It would be dubious for the first photographer to discover the aesthetic qualities of photography's golden hour to claim exclusive protection over photography in that aesthetic (it might fail as "method" or "idea"[390]), but it would be entirely laughable for photographers inspired by prior photos to make the same claim. Later photographers no more originate the technique than a historian authors the year of Napoleon's exile.

Bo Kim's recent work on scenes-a-faire re-explains the doctrine from a separate angle that emphasizes the non-authorial component of the work's value.[391] Apart from the matter of originality, he argues scenes-a-faire doctrine is distinctive because scenes-a-faire elements derive their value from audience expectations.[392] As applied to the definition of scenes-a-faire as elements "necessary" within a genre—a definition often criticized as vague—he argues the real question is whether the element is needed to satisfy audience expectations as set by "engage[ment] with existing popular and widely distributed creative works."[393]

The logic readily carries over to AI training. Machine learning may be directed toward expressive choices that are, quite literally, conventional: the macro lens effect of an Instagram-worthy food photo, or the pixelization of retro video-game art. A machine learning process might digest a series of caricature paintings to extract the various ways an artist can exaggerate a subject's ears, hair, and freckles. The individual elements are no doubt creative, perhaps even playful. Moreover, the use may be substitutive, depriving human caricaturists of commissions if training yields a model that can caricature photos to similar effect. But a significant portion of the value the artists lose will relate to satisfaction of artistic conventions that transcend any specific artist's authorial contributions.

Harder cases have arisen with newsworthy photos, where third-party interest drives a work's market value in a different way. Courts sometimes defer to such public interest to extend fair use. Consider *Time Inc. v. Bernard Geis*, a case dealing with a film of President Kennedy's assassination. Abraham

---

389 *See* Bo S.L. Kim, *Copyright's Public Reliance Interests*, 99 Wash. L. Rev. 107, 111 n.19 (2024).

390 *See* 17 U.S.C. § 102(b).

391 *See generally* Kim, *supra* note 389.

392 *Id.* at 146.

393 *Id.* at 147.

Zapruder, who filmed the event, "was by sheer happenstance at the scene taking home movie pictures with his camera."[394] When a later author used frames from the film in his book on the assassination, a court upheld the use as fair.[395] This outcome is easy to explain from the perspective of non-authorial value: the "public interest in having the fullest information available on the murder of President Kennedy"[396] was a source of non-authorial value that the court permitted the defendant to utilize. Zapruder contributed little that was authorial, having mainly been in the right place at the right time to capture the historic event.

The later decision in *Núñez v. Caribbean Int'l News Corp.* dealt with the unauthorized publication of nude photographs of a young woman who had been selected as Miss Puerto Rico.[397] The photographs had become newsworthy because of controversy over whether it was appropriate for someone who had posed for such pictures to be crowned with the pageant title.[398] A court upheld the publication as transformative use.[399] From the perspective of non-authorial value, the court endorsed the paper's exploitation of the value in the pictures as subjects of the news story, apart from any creativity or expression the photographer contributed.[400] Although the wholesale reproduction of the photographs also entailed reproduction of the photographer's creative choices, there was no authorial harm to speak of because the photographer established no injury to his photography business.[401]

In other news-photography cases, however, courts have sided with the copyright owner. Consider the more recent decision rejecting fair use in *McGucken v. Pub Ocean Ltd.*[402] Photographer Elliot McGucken captured a photograph of an

394 293 F. Supp. 130, 131 (S.D.N.Y. 1968).

395 *Id.* at 131–32, 146.

396 *Id.* at 146.

397 235 F.3d 18, 21 (1st Cir. 2000).

398 *See id.*

399 *Id.*

400 *See id.* at 22 ("[Defendant] sought not to 'scoop' appellant by publishing his photograph, but merely to provide news reporting to a hungry public.").

401 *See id.* at 25.

402 42 F.4th 1149 (9th Cir. 2022). The intervening case *Monge v. Maya Magazines, Inc.* denied fair use in a scenario involving photographs of a clandestine celebrity wedding. 688 F.3d 1164 (9th Cir. 2012). *Monge* distinguished *Núñez* as a case where the public was interested in the specific photo as a newsworthy object and rejected fair use in the case before it where the specific photos in question were not themselves the objects of public conversation. *Id.* at 1175.

ephemeral lake that appeared in Death Valley following heavy rain.[403] There is non-authorial value in such a photo because part of its splendor stems from the lake as a physical phenomenon. There is also no denying public curiosity, which could be satisfied, to a degree, by any photo of the lake if only the lake were still there to be photographed. The court nonetheless denied fair use for a magazine's unauthorized use.

So what distinguishes *McGucken* from *Núñez*? Two paths offer reconciliation. One is simply that, regardless of the magazine's right to exploit the work's non-authorial value, the harm to the photographer's market as a practical matter was too great. The very motivation for McGucken's photography was to license it to magazines and newspapers, and unauthorized use stood to destroy his primary market.[404] The other possibility is to integrate the strain of doctrine where courts have sometimes recognized authorship in the mere timing of a photo when the photographer has been deliberate about capturing the moment.[405] In the same vein, the *McGucken* court observed that the defendant had used the image primarily for its aesthetic impact rather than depiction of its factual subject matter—"it essentially use[d] the photos as visual 'filler.'"[406] Reproduction of a work in full for its aesthetic value is prone to usurp its authorial contributions, even if they are thin.[407] Though it may mean sometimes allocating photographers a portion of value they did not contribute, these observations provide a foothold to argue that using such an image without authorization usurps that which copyright assigns to the author.

D. Principles and Pragmatism

Taking exploitation of authorial value as our test, its application to AI training is sure to be contested. The cross-cutting lesson of the foregoing cases is that copying followed by market

403 *McGucken*, 42 F.4th at 1153.

404 *See id.* at 1163. *Monge* likewise cast doubt on uses that seek to "scoop" first publication of a story. *See* 688 F.3d at 1175–76.

405 *See* Hughes, *supra* note 375, at 380. This approach would leave room still to exclude Zapruder's fortuitous capture of tragic events in *Bernard Geis*. *See supra* notes 394–96 and accompanying text.

406 42 F.4th at 1159 (quotation omitted).

407 *See* Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 578–79 (1994) (disfavoring the use of a work "to get attention or to avoid the drudgery in working up something fresh"); *supra* note 199 (collecting cases where courts rejected fair use for derivative works that exploited the original works' intrinsic entertainment value).

harm is not fatal to fair use. Instead, courts allow competition when copying diverts value not attributable to the author. The photojournalism cases nonetheless show that deciding what value belongs to the author may not be as simple as filtering between protected and unprotected elements. As Justin Hughes observes, "the need for incentives, our concern for fairness, and our sense of beauty . . . may tend to make us look at the copyrightability of photographs generously."[408] To extrapolate, courts' assessment of copyright policy and the market realities of specific contexts shapes their determinations regarding the scope of authorship—and, by extension, the scope of authorial value.

The non-authorial value theory provides a principled framework for upholding AI training as fair use insofar as it exploits the value of facts, social conventions, and network effects rather than artists' expression. The capacity of AI systems to displace artists without copying their work suggests that, concretely, the harms many artists suffer stem from copying common styles or subject matters rather than appropriating the artists' distinctive authorial contributions.[409] Molly Shaffer Van Houweling's articulation of the "freedom to extract" unprotectable elements pushes the argument even further.[410] Her "focus on what the user *is* doing (extracting unprotected elements) as opposed to what the user is *not* doing (consuming the work's expression or exploiting its authorial value)"[411] may justify fair use in cases that are hard for the non-authorial value theory, like those where destruction of authorial value is unavoidable in the course of copying facts.[412] The authorial-value framework also encompasses judicial consideration of technical safeguards like de-duplication and filtering.[413] Practically speaking, these strategies matter because they diminish the likelihood that an AI system will reproduce the author's distinctive contributions to any given work.

---

408 Hughes, *supra* note 375, at 392. Jane Ginsburg once posited that, "[i]n copyright law, an 'idea' is not an epistemological concept, but a legal conclusion prompted by notions—often unarticulated and unproven—of appropriate competition." Jane C. Ginsburg, *No "Sweat"? Copyright and Other Protection of Works of Information after* Feist v. Rural Telephone, 92 COLUM. L. REV. 338, 346 (1992).

409 *See infra* subpart IV.B.

410 *See* Van Houweling, *supra* note 31, at 112.

411 *Id.*

412 *See id.* at 109, 122–25.

413 *See, e.g.*, Ayres & Balkin, *supra* note 345, at 9; *see also* Sag, *supra* note 9, at 338–43 (offering more detailed proposals).

Nonetheless, the non-authorial value theory and its application are not without challenges. The first U.S. Copyright Act's protection of "maps, Charts, And books,"[414] and its framing as "an act for the encouragement of learning,"[415] illustrates a historical willingness to protect fact-adjacent material. Carrying the argument forward, Ben Sobel questions whether the fact–expression dichotomy has merely been a "good-enough proxy" for identifying legitimate markets, inviting us to draw a new dividing line for AI.[416] *McGucken* also models an intuitive argument against AI training.[417] The court found that exploiting McGucken's photo for its aesthetic value encroached on that which belongs to the author;[418] one might similarly argue that an AI system exploits an author's contributions when it trains on a copyrighted image to replicate the image's aesthetics.

Reconciling these tensions requires closer examination of copyright policy and the consequences of copyright enforcement for AI developers and aggrieved parties. The road ahead is fraught, however, because copyright scholarship is ambivalent on whether the use of AI systems advances authorship and public discourse. Market realities further complicate the issue because enjoining AI training may not effectively redress harms to artists. The next Part takes up the problem of identifying and acting on copyright's goals in the face of these challenges.

IV
Copyright Cannot Solve Our AI Problems

Copyright provides a remarkably limited toolkit for dealing with the problems of AI—even the problems it poses for artists. Many artists facing displacement will have no legal basis to complain. The threat to most artists now comes from the creation of systems that can duplicate a range of common objects

---

414 *See* Benjamin L.W. Sobel, *On Copyright, "Facts," & Generative AI*, Digital Life Initiative (Sept. 23, 2024), https://www.dli.tech.cornell.edu/post/on-copyright-facts-generative-ai [https://perma.cc/VY8R-SVVD] (discussing the Copyright Act of 1790, Act of May 31, 1790, ch.15, 1 Stat. 124).

415 *Id.*; *see* Brauneis, *supra* note 9, at 39–40. Contemporary cases involving educational and journalistic fair use also support the point. Their value derives in large part from facts, third-party interest, or compulsory K–12 schooling, and the Copyright Act goes so far as to single them out as uses that are likely to be fair. *See* 17 U.S.C. § 107 (preamble). Yet courts have often denied fair use to protect the market for instructional texts and news. *See* 4 Nimmer & Nimmer, *supra* note 9, at § 13F.10[D][2].

416 *See* Sobel, *supra* note 414.

417 *See* McGucken v. Pub Ocean, 42 F.4th 1149 (9th Cir. 2022).

418 *See supra* note 406 and accompanying text.

and styles, not from copying these artists' specific creative contributions. These artists will face this threat even if system designers use training sets based on other artists' work. Copyright is not structurally equipped to address this problem because these artists lack the factual basis to assert a copyright claim. The following discussion situates copyright's limits and implications within the context of larger dilemmas for art and for tech concentration.

A. Our AI Problems

1. *Dilemmas in Copyright Policy*

Securing the future of art is fundamentally important to copyright policy. There is no consensus on how copyright should treat the fair-use question, however, because different accounts of copyright and AI art prioritize different implications. Support and opposition for fair use in this context breaks down largely across familiar lines. However, the distributive dimensions of the problem—where fair use may disadvantage individual, working artists relative to more financially secure rightsholders—may complicate the positions.

To start, copyright's basic concerns—whether fair use would facilitate the production of new works and access to them[419]—hinge on a set of empirical questions requiring further study. Production of new works may be curtailed to the degree that artists leave the field because generative AI puts them out of work.[420] Yet it may be accelerated to the degree that generative tools make it easier for people without traditional artistic skills to create new works[421] and popularize new artforms.[422] Generative AI's speed and ease of use will also reduce the costs of creating new works, expanding total output and potentially

[419] *See* Glynn S. Lunney, Jr., *Reexamining Copyright's Incentives-Access Paradigm*, 49 VAND. L. REV. 483, 485 (1996).

[420] *See* Dan L. Burk, *Cheap Creativity and What It Will Do*, 57 GA. L. REV. 1669, 1680–82 (2023).

[421] *See, e.g.*, Geddes, *supra* note 40, at 21. In the same spirit: "Generative AI can help people with disabilities create content that expresses their ideas, emotions, and perspectives, and can do so in different modalities and formats." Michael Mace, *The Upsides of Generative AI*, CONNECTED PROFESSOR (Spring 2023), https://connectedprof.iu.edu/articles/2023-spring/taking-note.html [https://perma.cc/SYF6-F5GU].

[422] *See, e.g.*, Tonio Inverness, Comment Letter on U.S. Copyright Office Notice of Inquiry on Artificial Intelligence and Copyright (Sept. 2023), https://www.regulations.gov/comment/COLC-2023-0006-2199 [https://perma.cc/3FCW-VCS9] (urging the Copyright Office to recognize the art of "synthography").

diminishing the need for copyright to motivate some forms of creative production.[423] The tradeoff is also contingent on the resolution of other copyright questions, such as the copyrightability of AI outputs.[424]

Then there are the deeper questions about copyright's role in society beyond the sheer number of works produced. Many prior accounts of copyright policy have emphasized that copyright's constitutional mandate, "[t]o promote the Progress of Science and useful Arts,"[425] is advanced by broad engagement in cultural production,[426] including in the realm of pop culture.[427] These arguments have historically been mobilized to advocate in favor of artists' free remixing or reworking of films, albums, and other works from major media conglomerates.[428] Similar arguments would seem to apply to generative AI, which allows for its own unique forms of reimagining and resplicing culture.[429]

These arguments have historically been bolstered, however, by a Robin-Hood dynamic whereby the relaxation of copyright takes from the rich (major media corporations) and gives to the poor (the aspiring artist or even the fan-fiction writer).[430] With generative art, the cost is more keenly felt by individual artists, whose works may be used to train the AI models that displace them.[431] Meanwhile, critics observe that the

---

423 *See* Burk, *supra* note 420, at 1680.

424 *See generally Copyright Registration Guidance: Works Containing Material Generated by AI*, 88 Fed. Reg. 16190, 16192 (Mar. 16, 2023) (stating the Copyright Office's position).

425 U.S. Const. art. I, § 8, cl. 8.

426 *See* Jack M. Balkin, *Digital Speech and Democratic Culture: A Theory of Freedom of Expression for the Information Society*, 79 N.Y.U. L. Rev. 1, 39; Rebecca Tushnet, *Copy This Essay: How Fair Use Doctrine Harms Free Speech and How Copying Serves It*, 535 Yale L.J. 535, 566–67 (2004) ("Copying can serve as self-expression . . . and it can work as affirmation, a way of connecting to a larger group.").

427 *See* Anupam Chander & Madhavi Sunder, *Everyone's a Superhero: A Cultural Theory of "Mary Sue" Fan Fiction as Fair Use*, 95 Calif. L. Rev. 597, 599–600 (2007).

428 *See generally* Lawrence Lessig, Remix: Making Art and Commerce Thrive in the Hybrid Economy (2009).

429 *See supra* note 184 and accompanying text (introducing Katrina Geddes's account of semiotic democracy). The interest in promoting broad access to pop culture would also seem to counsel in favor of relaxing character copyright in this space to avoid the front-end or back-end expressive burdens associated with the Spider-Man Problem. *See supra* section I.D.3.

430 Sobel, *supra* note 9, at 86–89.

431 *See infra* section IV.A.2.

highest-profile beneficiaries of fair use may be parties like Google and Microsoft-funded OpenAI.[432]

Although copyright policy is not necessarily concerned with distributive outcomes for their own sake, the concentration of revenues among a smaller number of players may reduce the number of speakers and the diversity of creative expression. This contraction in the range of expression would impede copyright's goal of democratizing speech and cultural production.[433] Neil Netanel's writing on the importance of copyright to support the existence of a free and independent press is particularly pertinent,[434] especially if we shift briefly from visual art to consider the New York Times' suit against OpenAI.[435] Regardless of how copyright treats AI training as a general matter, we might consider special regimes to protect journalism, much as the Supreme Court once fashioned a bespoke quasi-property rule to protect the news of the day in *International News Service v. Associated Press*.[436]

Then there is the question of copyright's role in shaping competition and technology generally. In broad strokes, copyright has paved the way for new technologies that allow for new exploitations of existing works. Courts have rejected suits in which copyright owners attempted to leverage their rights to squelch a new market, rather than enter it,[437] across contexts as different as the player piano,[438] cable television,[439] the VCR,[440] and MP3 players.[441] I have previously argued these

432 Sobel, *supra* note 41, at 32 ("[T]here are strong doctrinal reasons to doubt that a rising tide for AI would raise all boats."). Others raise the contrary argument that employing copyright to restrict training may favor intermediaries rather than creators. *See* Craig, *supra* note 9, at 24–26. Although there may be truth to both positions, the possibility that denial of fair use may favor established platforms lends credence to the latter. *See infra* subpart IV.C.

433 *See supra* notes 426–27 and accompanying text.

434 *See* Neil Weinstock Netanel, *Copyright and a Democratic Civil Society*, 106 YALE L.J. 283, 352–62 (1996).

435 Complaint, N.Y. Times Co. v. Microsoft Corp., No. 1:23-cv-11195 (S.D.N.Y. filed Dec. 27, 2023).

436 Int'l News Serv. v. Associated Press, 248 U.S. 215, 236 (1918).

437 Ginsburg, *supra* note 10, at 1617.

438 *See* White-Smith Music Publ'g Co. v. Apollo Co., 209 U.S. 1 (1908).

439 *See* Fortnightly Corp. v. United Artists Television, Inc. 392 U.S. 390 (1968); *see also* Teleprompter Corp. v. Columbia Broad. Sys., 415 U.S. 394 (1974) (extending this holding when copyright owners later attempted to participate in the emerging market).

440 *See* Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417 (1984).

441 *See* Recording Indus. Ass'n of Am. v. Diamond Multimedia Sys., Inc., 180 F.3d 1072 (9th Cir. 1999).

decisions advanced policies internal to copyright with respect to public engagement with works and the promotion of more diverse media;[442] others have arrived at similar conclusions by arguing these decisions used copyright as a tool to advance policy objectives in communications or innovation.[443]

The rub with extending that line of thinking to generative AI, however, is that the prior scenarios lent themselves to solutions where all players could benefit. Those technologies involved the use of existing works in their entireties, and it was intuitive to monetize the exploitation of any given work. Cable television providers, for example, could pay broadcasters for retransmission of their programs, and movie studios could monetize film catalogs by releasing films on videocassettes.[444] In the abstract, it is quite plausible that the economic gains of generative art systems will be greater than the losses, but the arcane nature of the training and design makes it difficult if not impossible to trace outputs back to specific training works to determine who should be paid. System designers might bypass the attribution problem by training AI models on a narrower set of training works contributed by only a handful of artists and dividing credit among them. But this possibility accentuates another problem: a capable generative art system may compete with the works of artists whose works were not included in the training data.[445] Any solution premised on rewarding only those artists whose specific works were copied—whether through litigation or royalty schemes—does nothing for this broader constituency.

2. *Popular Account*

In public debate, the problem posed for artists is more straightforward. Generative art is poised to flood the market with remarkably cheap, quick, and customizable works. All three features stand to undercut the market for existing works and to divert new commissions. Cheapness has obvious ramifications: industry representatives fear that, even though AI works may be lower quality, people will often be willing to

---

442 BJ Ard, *Taking Access Seriously*, 8 TEX. A&M L. REV. 225, 229–30 (2021).

443 *See, e.g.*, von Lohmann, *supra* note 11 (innovation policy); Timothy Wu, *Copyright's Communications Policy*, 103 MICH. L. REV. 278 (2004) (communications, innovation, and competition policy).

444 Ard, *supra* note 442, at 248.

445 *See infra* section IV.A.2.

settle.[446] Indeed, the fact that AI outputs are so cheap, with such fast turnaround, means that a consumer loses little by trying a few test prompts before deciding whether to utilize AI or instead commission a human artist. Customizability also adds unique appeal to generative art. Consumers may prefer images specifically tailored to their demands over generic stock photos and clip art. Generative art may likewise compete with bespoke categories of art that historically required commissions—for example, an AI could easily be trained to create caricatures from uploaded photos or paint oil paintings imagining the family pet as lord or royal.

Compounding the problem, the result may be to spare famous artists and pile harm on the lesser known. Famous artists, living and dead, trade on more than the raw aesthetic appeal of their works: they also benefit from the cachet of their established names and demand among art aficionados for authentic works linked to famous creators.[447] Art aficionados will still pay a premium for famous artists' works even if AI can create aesthetically comparable images. Meanwhile, the artists who create stock images and magazine illustrations, and those who work behind the scenes on film and video-game animations, compete primarily on visual appeal. We can easily imagine generative AI systems replacing them or driving their prices down by providing works that are "good enough" and much cheaper. Some people may specifically choose to support human artists, just as some choose organic or fair-trade goods over cheaper options, but popular sympathies are weak predictors of market behavior.[448]

Individual artists may also face disadvantages relative to larger rightsholders owing to the expansiveness of character copyright. The copyrighted characters who headline major films, video games, and television programs enjoy broad

---

446 Copyright Alliance, *supra* note 6, at 95.

447 Indeed, the rise of non-fungible tokens ("NFTs")—with their promise of establishing authenticity and genuine connection with respect to digital art—is partly explained by this demand. *See* Amy Adler, *Artificial Authenticity*, 98 N.Y.U. L. REV. 706, 760–61 (2023). The fall of NFTs likewise resulted from realization that this claim, among many others, was wildly overblown. Thomas D. Haley, *Embracing Digital*, 101 N.C. L. REV. 619, 632 (2023).

448 *But see* Jacob Noti-Victor, *Regulating Hidden AI Authorship*, 111 VA. L. REV. (forthcoming 2025) (July 29, 2024 draft on file with author) (advocating for AI transparency to facilitate consumers' choice between AI and non-AI works). For further exploration of the intersection between law and consumer "preferences for processes," see generally Douglas A. Kysar, *Preferences for Processes: The Process/Product Distinction and the Regulation of Consumer Choice*, 118 HARV. L. REV. 526 (2004).

protection.[449] Rightsholders like Mattel thus stand in the position to enjoin the use of AI-generated Barbie images regardless of whether it is fair to use Barbie images in training.[450] Not so for the nature photographer or the impressionistic painter, whose subject matters and techniques remain fair game. Even a budding comic book artist might find her characters unprotected relative to those of major publishers, in light of scholarship observing that courts treat market success as a proxy for whether a character is recognizable enough to command copyright protection.[451]

Generative art thus stands to displace creative workers just as AI and automation threaten other employment sectors.[452] Artists are in a unique position, however, because copyright gives them a potential legal tool. If roboticists study and copy the movements of factory workers to program robots that eliminate manufacturing jobs, the workers have no proprietary claim to enjoin studies of on-the-job performance. When AI designers copy the work of human artists, however, the artists can plausibly invoke copyright to argue their own creative output has been unlawfully used against them.

B. Inefficacy of Copyright

Whatever our assessment of the right policy for art, copyright is poorly suited to addressing the harms that flow from generative art systems. Blake Reid frames the problem in terms of the "structural limits" of copyright—its inability to go beyond binary decisions of liability or fair use, and even then only in cases involving unauthorized use.[453] Oren Bracha finds copyright's tools for regulating the copying of discrete works ill-suited to address AI challenges that involve not discrete works, but instead the regulation of aggregate metainformation.[454]

We can see these limitations concretely in copyright's inability to guard against the economic harms inflicted by systems trained on licensed works. Copyright seemed efficacious

---

449 *See supra* section I.D.3.

450 5 NIMMER & NIMMER, *supra* note 9, § 20.05[C][2][b].

451 *See* Shani Shisha, *Commercializing Copyright*, 65 B.C. L. REV. 443, 467–69 (2024).

452 Sobel, *supra* note 9, at 82.

453 *See* Reid, *supra* note 10, at 54–55 ("[C]opyright's structural limits both undermine copyright's capacity to serve as a pluralistic public governance regime and disempower stakeholders who hope that copyright will vindicate their interests.").

454 Bracha, *supra* note 9, at 223.

at the time various plaintiffs first sued Stable Diffusion, when the major image-generation systems—DALL-E 2, Midjourney, and Stable Diffusion—all relied heavily on unlicensed works scraped from the internet.[455] Even then, however, computer scientists could imagine one day using licensed or public-domain works to build a comparable system.[456]

That possibility was soon realized when Adobe's Firefly system, trained on licensed images, launched in June 2023.[457] Getty Images, itself a plaintiff in one of the suits against Stable Diffusion,[458] followed shortly thereafter with its own system that it advertises as "commercially safe" because it excluded unlicensed works from the training set.[459] Both systems function as diffusion-based image generators. They differ from prior systems primarily in that they were trained on images already licensed or owned: Adobe, as a graphics giant, and Getty Images, as the leading name in stock photos, had already stockpiled high-quality images that they could use to train AI models without need to scrape additional images from the internet.[460]

Commercially safe systems pose the same fundamental threat to artists. Even accepting that models trained on the wider internet are better because their training sets are more comprehensive, the commercially safe systems are already sufficiently advanced to produce cheap, quick, and customizable

455 *See supra* notes 3–4 and accompanying text.

456 *See, e.g.*, Amanda Levendowski, *How Copyright Law Can Fix Artificial Intelligence's Implicit Bias Problem*, 93 WASH. L. REV. 579, 614–16 (2018) (discussing use of public domain for training AI systems); Aaron Gokaslan et al., *CommonCanvas: An Open Diffusion Model Trained with Creative-Commons Images*, ARXIV (Oct. 25, 2023), https://arxiv.org/pdf/2310.16825 [https://perma.cc/X47K-35UP] (examining feasibility of constructing a system using Creative Commons works). To be sure, Levendowski's central argument was that using a limited pool of works would lead to poor AI models because it would exacerbate bias, Levendowski, *supra*, at 614–19, and others have posited that the costs of licensing would be insurmountable, *see* Lemley & Casey, *supra* note 31, at 748. The developments discussed in this subpart and the next nonetheless demonstrate the commercial viability of training generative image systems without relying on fair use. The technology and market dynamics may vary for other classes of systems. Text systems, for example, may need access to specific works because fidelity to the facts therein is necessary to make text generators useful and non-defamatory.

457 Shrivastava, *supra* note 44.

458 *See* Complaint, Getty Images (US), Inc. v. Stability AI, Inc., No. 1:23-cv-00135-UNA (D. Del. filed Feb. 3, 2023).

459 *See* Getty Press Release, *supra* note 44. Facebook parent-company Meta launched its own generative art system trained on images licensed from users later the same year, Edwards, *supra* note 4, but we will take it up separately because training on user data presents distinct issues, *see infra* subpart IV.C.

460 *See* sources cited *supra* notes 457 & 459.

images that are adequate to serve many users' needs.[461] Moreover, competitive pressure from these systems underscores the point that the threat to artists often stems from depicting features beyond specific artists' protected expression.[462] Consider artists who produce stock photos, or photographers who create personalized, on-site images for commercial clients. A real estate broker might wish to create an advertisement featuring a diverse array of happy clients. Prior to the advent of AI, the broker might license a suitably generic stock photo, or it might splurge to hire a photographer who could photograph clients (or models) at its properties. But Adobe has introduced "generative fill," a tool that allows users to erase part of an image and enter a prompt asking for the insertion of particular people or things in a style that matches the rest of the photo.[463] The broker could take photos at its properties, erase a few person-sized rectangles, and ask the system to manufacture happy, smiling people to fill the spaces. This sort of activity may cost the stock photo provider and the photographer business.[464] However, the losses do not follow from an AI creator copying their specific expression, much less using it against them. Instead, these artists face problems because the systems can produce common subject matter in conventional styles at a fraction of the customary time and expense.

And the use of existing image stockpiles is not the only path for creating AI systems without relying on fair use. Other nations have adopted express copyright exemptions for machine learning, some of which may extend to AI training.[465] In addition, researchers have also proposed using existing AI

---

461 *See, e.g.*, Mark Hachman, *The Best AI Art Generators: Bring Your Wildest Dreams to Life*, PCWorld (Apr. 4, 2023), https://www.pcworld.com/article/1672975/the-best-ai-art-generators-for-you-midjourney-bing-and-more.html [https://perma.cc/4KXX-Z5RZ] ("Adobe's model doesn't seem as creatively free as some others, but makes up for it with its slick, professional look."); *see also supra* note 446 and accompanying text (on users' willingness to settle).

462 *See supra* section I.D.2.

463 *See* Jess Weatherbed, *Adobe Is Adding AI Image Generator Firefly to Photoshop*, Verge (May 23, 2023), https://www.theverge.com/2023/5/23/23734027/adobe-photoshop-generative-fill-ai-image-generator-firefly [https://perma.cc/P4SS-N54T].

464 *See* Derick David, *How AI Is Killing the Stock Photo Industry*, Medium (July 6, 2023), https://medium.com/utopian/how-ai-is-killing-the-stock-photo-industry-b41d3b4ba8ae [https://perma.cc/5GZW-TGFA].

465 *See* Matthew Sag & Peter K. Yu, *The Globalization of Copyright Exceptions for AI Training*, 74 Emory L.J. (forthcoming 2025) (manuscript at 16–21) (Oct. 4, 2024 draft on file with author) (detailing express exceptions in Japan, the United Kingdom, the European Union, and Singapore).

systems to generate reams of synthetic data for future training: "AI systems can generate synthetic data, which then trains other AI systems."[466] Although the legal theory is untested, it is plausible that these methods could train on unlicensed images but then create millions of new outputs on which a subsequent model could train, and repeat the process as needed to create enough degrees of separation to eliminate infringement claims as to the original training images.[467]

The viability of systems that do not rely on scraping the internet for unlicensed training data indicates that fair use is not the live-or-die question for generative AI. But the possibility of creating licensed systems does not render fair use superfluous. As the next subpart details, the decision still carries deep consequences in determining who creates, deploys, and controls these systems. Indeed, for those arguing that denial of fair use will combat tech monopolization, the move may backfire.

C. Generative Art and Tech Concentration

Fair use will also bear on competition in the AI sector. The FTC has urged using copyright to enjoin the use of unlicensed training data.[468] They argue the move would be pro-competitive because it would stop dominant tech firms from scraping training data from the internet and cornering the nascent AI market.[469] The theory is dubious:[470] the absence of evidence that denying fair use will promote competition is compounded by the absence of evidence it will assist artists.[471] Worse, this approach may backfire with respect to competition. Raising the

---

466 *See* Lee, *supra* note 44, at 36.

467 *See id.* at 23–24.

468 *See* U.S. Federal Trade Commission, *supra* note 8, at 5–6.

469 *See id.*

470 Scholars have criticized the FTC submission for both its opacity and its substance: "the FTC's submission is not a model of clarity," and fails to "explicitly refer to or analyze" the relevant caselaw. Pamela Samuelson, Christopher Jon Sprigman & Matthew Sag, Reply Comments on U.S. Copyright Office Notice of Inquiry on Artificial Intelligence and Copyright, 2–3 (Dec. 6, 2023), https://www.regulations.gov/comment/COLC-2023-0006-10299 [https://perma.cc/2C7J-D4RA]. Daryl Lim and Peter Yu raise the specific objection that the FTC's approach is not tailored to the nature of the problem, noting "we are struggling to address a paradox of bigness: substantial size is necessary for AI firms to thrive in large-language models (LLMs) or big data analytics, yet the firms' gigantic size and ever-growing market power may pose significant harm to competition and consumer interests." Daryl Lim & Peter K. Yu, *The Antitrust–Copyright Interface in the Age of Generative Artificial Intelligence*, 74 EMORY L.J. (forthcoming 2025) (manuscript at 11) (Mar. 3, 2024 draft on file with author).

471 *See* Craig, *supra* note 9, at 2.

barriers to obtaining training data may instead entrench established players, particularly the social media platforms that are positioned to train AI models on their established hoards of user data.

These dynamics stem only partly from incumbents' greater resources. Some niche players already pursued business models that coincidentally endowed them with a large supply of training works: Getty Images with the stock photos it has amassed for direct licensing and Adobe with the images that it has licensed or otherwise acquired in connection with its graphics software packages.[472] Much larger enterprises, like Microsoft or Google's parent company Alphabet, presumably have the resources to acquire similar materials.[473]

The other part of the story has to do with the advantages digital intermediaries have won over nearly three decades of cyberlaw jurisprudence. Those who are opposed to the use of their materials as training data have two primary weapons: copyright and privacy law. As case in point, in addition to the copyright claims explored above, plaintiffs have also asserted privacy claims against Google and OpenAI for scraping their online materials.[474] These weapons are blunted, however, because legal developments around terms of service allow platforms to invoke boilerplate as a shield. As to copyright, platforms can simply incorporate language whereby users grant a license to use their uploaded content.[475] Privacy operates much the same way. Many privacy claims stem from a platform's terms of service; others are statutory. Typically, however, both classes of privacy claims are waived through users' acceptance of terms of service.[476] This explains why companies like Meta—parent company of Facebook and Instagram—can proceed with the creation of their own image-generation systems using

[472] *See supra* notes 457–59 and accompanying text.

[473] *But see* Lemley & Casey, *supra* note 31 (predicting that denying fair use would make it prohibitively difficult to train generative AI).

[474] *See* Complaint, J.L. v. Alphabet, Inc., No. 3:23-cv-03440 at 1 (N.D. Cal. filed July 11, 2023) (alleging "mass theft of personal information"); P.M. v. OpenAI LP, No. 3:23-cv-03199 at 1–2 (N.D. Cal. filed June 28, 2023) (alleging OpenAI has "stolen and misappropriated personal information at scale").

[475] *See* Amit Elazari Bar On, *Unconscionability 2.0 and the IP Boilerplate: A Revised Doctrine of Unconscionability for the Information Age*, 34 BERKELEY TECH. L.J. 567, 610–11 (2019).

[476] Omer Tene, *Privacy: The New Generations*, 1 INT'L DATA PRIVACY L. 15, 25–26 (2011) ("In addition, the promises made in a privacy policy often ring hollow given the right reserved by service providers to unilaterally modify or amend its terms at will.").

user-uploaded content with (relative) impunity to copyright and privacy claims.[477] These advantages place established platforms ahead in the AI space.

These dynamics do not lend themselves to easy resolution of the fair-use question, much less to solutions for tech monopolization. We should not reject fair use thinking that this rejection would somehow promote competition. In the same spirit, we should be cautious about denying fair use as a matter of copyright policy without first weighing what most artists stand to gain (little)[478] relative to how much this move may entrench the privileged position of some established platforms (perhaps substantially). Making headway on these issues requires us to acknowledge generative art is not just or even primarily a copyright issue—it also poses urgent questions for competition law and platform regulation alongside its challenges for the future of employment.[479]

Conclusion

Generative AI poses significant challenges for copyright doctrine and copyright's normative foundations. This Article provides a guide to navigating these challenges, starting with a technical primer on AI training and system design. Understanding the multi-phase nature of AI development helps clarify why drawing a direct line from the purpose of copying during training to the goals of the final AI system is untenable. By highlighting the complications that generative AI poses for questions of purpose, market impact, and use or non-use of expression, this analysis also exposes the limits of prevailing approaches to fair use.

The non-authorial value framework offers a new approach. It asserts that market harm alone is not the question—copyright has historically punished diversion of value contributed by

---

[477] *See* Reece Rogers, *Facebook Trains Its AI on Your Data. Opting Out May Be Futile*, Wired (Sept. 7, 2023), https://www.wired.com/story/facebook-trains-ai-your-data-opt-out [https://perma.cc/2ELG-YR7Q]. Meta will inevitably face difficulties because users appear to upload images they do not own, hence Meta AI's over-eager Spider-Man depiction in *supra* Figure 13.

[478] *See supra* section IV.A.2.

[479] For thoughtful treatment of AI-competition law beyond copyright, see Lim & Yu, *supra* note 470, at 11 (addressing the "paradox of size" where substantial size is necessary for AI development); Solow-Niederman, *supra* note 92, at 688–90 (exploring public provision of the necessary algorithms and data sets); and Tejas N. Narechania, *Machine Learning as Natural Monopoly*, 107 Iowa L. Rev. 1543, 1557 (2022) (exploring the prospect of regulating AI as a natural monopoly when its high fixed and operational costs bar market entry).

an author while leaving room for copyists to exploit other facets of a work.[480] Some forms of AI training may qualify as fair under this test insofar as they exploit the value of styles and subject matters that authors cannot monopolize. Courts nonetheless face hard questions where the exploitation of seemingly non-authorial value imperils wider fields of creative endeavor. Concern for the future of art might push courts to define the scope of authorship broadly in this context, narrowing fair use.

These impulses must be tempered, however, by recognition that refining fair use is insufficient to address generative AI's potential impacts. Copyright alone cannot address the economic pressures AI places on artists, especially those who face competition from systems that do not actually copy their works. Moving forward requires going beyond fair-use debates to consider a wider array of policy tools—including antitrust, platform regulation, labor protections, and social welfare initiatives—to protect creators and to chart an equitable future for creative production.

480 More broadly, this dissection of types of market harm highlights the normative questions that remain unresolved if we take seriously the Court's embrace of market logic in *Warhol*. Andy Warhol Found. v. Goldsmith, 598 U.S. 508, 528 (2023) (designating "the problem of substitution" as "copyright's bête noire").