ARTificial: Why Copyright Is Not the Right Policy Tool to Deal with Generative AI

Micaela Mantegna

ABSTRACT. The rapid advancement and widespread application of Generative Artificial Intelligence (GAI) raise complex issues regarding authorship, originality, and the ethical use of copyrighted materials for AI training.

As attempts to regulate AI proliferate, this Essay proposes a taxonomy of reasons, from the perspective of creatives and society alike, that explain why copyright law is ill-equipped to handle the nuances of AI-generated content.

Originally designed to incentivize creativity, copyright doctrine has been expanded in scope to cover new technological mediums. This expansion has proven to increase the complexity and uncertainty of copyright doctrine's application—ironically leading to the stifling of innovation. In this Essay, I warn that further attempts to expand the interpretation of copyright doctrine to accommodate the particularities of GAI might well worsen that problem, all while failing to fulfill copyright's stated goal of protecting creators' rights to consent, attribution, and compensation.

Moreover, I argue that, in that expansion, there is the peril of overreaching copyright laws that will negatively impact society and the development of ethical AI. This Essay explores the philosophical, legal, and practical dimensions of these challenges in four parts.

INTRODUCTION

With the rapid evolution of Generative Artificial Intelligence (GAI), what once were theoretical questions about artificial creativity became a matter of reality for users, as well as a heated litigation battleground for companies. As some of these Artificial Intelligence (AI) systems were trained with copyright-protected works, ethical and legal questions arose:

- How should authors be compensated, and under which circumstances?

- Is training with unlicensed works a fair use, or an infringing one?
- Are the outputs generated by GAI original or derivative works?
- Should the outputs be entitled to some form of copyright protection? If so, how should we deal with AI authorship?
- Should we distinguish between autonomous and assisted AI creations? Can AI be assigned rights as an author for its contributions in some capacity? If so, where should the line be drawn? Or should all AI creations be in the public domain?
- Would assigning copyright to AI works create an imbalance in the public domain due to the massive scale and velocity at which AI can operate?
- Should the law treat differently the outputs of AI versus those of human authors? How would we justify such a framework without affecting the principle of "aesthetic neutrality"? How does that fare with the definition of "creativity" as a requisite for copyrights law's "originality" requirement?

Answering those questions in terms of copyright requires delving into definitions of creativity, originality, and art appreciation, as well as questioning philosophical rationales that justify the existence of intellectual property (IP). IP is no longer a niche area of law designed to regulate relationships between authors and publishers. It has become ubiquitous in our digital society. Everyone is now potentially a creator under copyright law, whether they are posting a photo on social media or tweeting a short haiku. Our smart devices, from cars to pacemakers and everything in between, are powered by software protected by copyright and digital-rights management. The media we consume in digital form is regulated by a complex landscape of contracts, Terms of Service, and IP regulations. Consequently, we have in recent decades witnessed an expansion of IP law. The definitions of core concepts and principles have been stretched to fit new scenarios and domains that transcend the doctrine's initial scope—sometimes to the point of thwarting its founding purpose. Paradoxically, while initially having the goal of promoting science and arts, copyright doctrine, as expanded to new scenarios and domains, has in many ways limited the cultural impact of works and the preservation of and access to knowledge.[1] The rapid expansion of the

---

1. *See* Siva Vaidhyanathan, Copyrights and Copywrongs: The Rise of Intellectual Property and How It Threatens Creativity 12-16, 183-89 (2001).

Internet has only heightened these effects,[2] leading to what is often referred to as the "copyright wars."[3]

Further, answering the above questions in terms of AI policy requires understanding AI in the context of ethics, economics, and culture, as well as AI's deployment in a digital society. As a technology, AI's implementation triggers different legal fields related to innovation, such as data protection, consumer protection, and antitrust. Therefore, a holistic policy solution to the GAI problem cannot be articulated just by thinking from the copyright corner.

As I have already explored those questions related to IP and AI in depth elsewhere,[4] this Essay tackles a different facet of this complex problem. In this Essay, I argue that the copyright framework might not be the right policy tool to deal with the consequences of GAI. To that end, I explore the multiple ways that AI impacts society, labor, and innovation, while highlighting issues of consent, attribution, and fair compensation for artists.

This Essay proceeds in four Parts. In Part I, I contextualize the policy debate by providing definitions of relevant terms from the field of AI and GAI, surveying some technical considerations about how AI technologies work. Next, I lay the groundwork for this Essay's argument in Part II by exploring key concepts surrounding creativity, ethics, and copyright. Then, I present in Part III an overview of the ethical concerns surrounding GAI, challenging the conflation of unethical behavior with illegality in public discourse that has been fueled by the intense cultural conflict between artistic communities and AI users. I argue that the impact of GAI cannot be framed just in terms of copyright law, isolated from the broader context of AI ethics. Specifically, I explore some of the societal harms of AI, including threats to democracy, perpetuation of inequality, emotional manipulation through AI companions, displacement of creative labor, and exploitation of data workers in the Global South, which can be framed as a form of automated colonialism.

Finally, in Part IV, I introduce some initial arguments to consider why copyright law, as a policy tool, is not well equipped to provide answers to the

---

2. *See* CORY DOCTOROW, NEIL GAIMAN & AMANDA PALMER, INFORMATION DOESN'T WANT TO BE FREE: LAWS FOR THE INTERNET AGE 84-131 (2014).

3. *See* DAVID BOLLIER, BRAND NAME BULLIES: THE QUEST TO OWN AND CONTROL CULTURE 145 (2008). *See generally* PETER BALDWIN, THE COPYRIGHT WARS: THREE CENTURIES OF TRANS-ATLANTIC BATTLE (2014) (exploring the historic copyright wars that underlie those of the current day).

4. *See generally* MICAELA MANTEGNA, ARTEFICIAL: CREATIVIDAD, INTELIGENCIA ARTIFICIAL Y DERECHO DE AUTOR (2022) (exploring different considerations at the intersection of Artificial Intelligence (AI), intellectual property (IP), and copyright by introducing a taxonomy of different elements that are relevant for legal analysis of Generative Artificial Intelligence (GAI) models and exploring the challenges AI poses to copyright doctrine in terms of authorship, originality, infringement, and the public domain).

questions raised by the use of GAI, both from the perspective of creatives and society at large. To that end, I consider how copyright expansion could affect the whole ecosystem of AI innovation, impact ethical principles of AI development, and further skew the balance between IP rights and culture in the digital context. I conclude that trying to fit the tension created by GAI into the copyright framework can put at risk already strained concepts like fair use, incentivize a high-risk landscape that can cannibalize creative industries (something I call "ouroboros copyright"), and ultimately have a chilling effect on the whole Internet ecosystem.

These arguments are meant to present a cautionary tale of unintended consequences. IP is a strange field in which the law attempts to apply a uniform framework to protect a multimillion dollar movie just as much as a casual drawing. And while legal protection might aspirationally be in place for both, the power dynamics of enforcing copyright protection side with those with deep pockets.

A survey of current AI copyright litigation predominantly reveals big players on both sides, representing both Big Tech and legacy creative intermediaries.[5] The way these forces are aligning and some of the arguments made resemble the power dynamics and tensions observed in past landmark copyright litigation that was instrumental to shaping our current Internet and digital landscape.[6] Below the surface of the "artist versus AI" dilemma, there is a reenactment of historic Internet-governance battles and a reappearance of old arguments (like the legality of linking, or web scraping) coming from the "usual suspects."

Creatives are used by other players to legitimize these old debates and reframe the core struggle as a noble fight. The narrative has been cleverly framed as an ethical battle for artists' rights against AI, but beyond what meets the eye, the structure of creative markets and the way copyright law and institutions work in practice might not guarantee that money will be redistributed after being captured by the usual intermediaries. Applying the copyright framework is by no means a guarantee that artists will be adequately compensated, and worse, the expansion of copyright may lead to the introduction of an unfair stratification of the labor that goes into AI training. Such new divisions could result in the unequal treatment of workers involved in the AI supply chain, creating a disparity between intellectually creative labor and other forms of labor essential for AI training, like content labeling or moderation.

---

5. *See AI Litigation Database – Search*, ETHICAL TECH INITIATIVE D.C. GEO. WASH. L. SCH., https://blogs.gwu.edu/law-eti [https://perma.cc/77YT-ENNX].

6. *See, e.g.*, Google LLC v. Oracle Am., Inc., 141 S. Ct. 1183 (2021) (deciding whether Google's copying of a portion of the Java programming language constituted fair use); Authors Guild v. Google, Inc., 804 F.3d 202 (2d Cir. 2015) (deciding whether Google's copying of authors' copyrighted works onto its Google Books and other platforms constituted fair use).

Two caveats warrant mention. First, I do not argue in this Essay that the labor that goes into AI training should remain uncompensated. There is no doubt that AI systems are threatening the livelihood of creative workers by unethically appropriating their work as training fuel to then mimic their styles, voices, or performances. AI is extracting creative inferences, modeling creativity, and in doing so, reshaping labor. This is another point where GAI connects with AI ethics at large. Back in 2015, Professor Jerry Kaplan predicted this asymmetry created by AI automation and its effects on labor, leading to a polarization of wealth.[7] The asymmetry reproduces a classic capitalistic dynamic, in which those who own the means of production are in a better position to accumulate long-term wealth and set the rules to defend it, even from those who have a rightful claim to participate in that wealth. Copyright law is no exception to this rule, and litigation has proven a useful tool for those with pockets deep enough to play the long game.[8]

But such litigation is hard to sustain for individual artists, as it is expensive, extremely complex, time consuming, and garners unpredictable results.[9] What's more, litigation has geographical limitations in terms of jurisdictions and applicable law.[10] Collective litigation and class actions pose other risks: people entitled to compensation might fail to pursue it,[11] and members of the class have little

---

7. *See* JERRY KAPLAN, HUMANS NEED NOT APPLY: A GUIDE TO WEALTH AND WORK IN THE AGE OF ARTIFICIAL INTELLIGENCE 3-16 (2015).

8. *See, e.g.*, Carey Dunne, *More than 40 Artists and Designers Accuse Zara of Plagiarism*, HYPERALLERGIC (July 29, 2016), https://hyperallergic.com/314625/more-than-40-artists-and-designers-accuse-zara-of-plagiarism [https://perma.cc/H46Q-V62F]; Gabby Bess, *How Fashion Brands Like Zara Can Get Away with Stealing Artists' Designs*, VICE (July 21, 2016, 4:35 PM), https://www.vice.com/en/article/nejwdz/how-fashion-brands-like-zara-can-get-away-with-stealing-artists-designs-tuesday-bassen [https://perma.cc/DZ8T-F7HG]. This phenomenon is more prevalent in the fashion industry in the United States due to the particularities of copyright law. *See generally* Kal Raustiala & Christopher Jon Sprigman, *The Piracy Paradox: Innovation and Intellectual Property in Fashion Design*, 92 VA. L. REV. 1687 (2006) (examining the apparel industry's widespread practice of design copying).

9. *See* Terrica Carrington, *A Small Claims Court Is on the Horizon for Creators*, COPYRIGHT ALL. (Oct. 4, 2017), https://copyrightalliance.org/small-claims-court-on-the-horizon [https://perma.cc/6NGB-ZKU8] ("According to an AIPLA report on the costs associated with IP litigation, the average cost of litigating a copyright infringement case in federal court from pretrial through the appeals process is $278,000. To put that number in perspective, on average, a *full-time* book author made only $17,500 from writing in 2015.").

10. For an illustration of some of these difficulties, see Joe Skrebels, *Game Developer Accuses Real-Life Weapons Manufacturer of Stealing Its Gun Design . . . Twice*, IGN (Apr. 20, 2022, 11:22 AM), https://www.ign.com/articles/stolen-gun-kalashnikov-oceanic-mp-155-ultima-ward-b [https://perma.cc/EW89-QTBV].

11. *See* Deborah R. Hensler, *Justice for the Masses? Aggregate Litigation and Its Alternatives*, 143 DAEDALUS 73, 74 (2014); Andrew J. Pincus, *Unstable Foundation: Our Broken Class Action System and How to Fix It*, U.S. CHAMBER INST. FOR LEGAL REFORM 3, 8 (Oct. 2017),

control over the litigation strategy, are precluded from pursuing separate individual litigation, or are eventually barred from pursuing litigation on account of the doctrine of res judicata and estoppel.[12] This Essay reflects on the motives and rationales that make copyright law an inadequate framework to address compensation for the use of copyrighted works as training materials for GAI. Possible solutions such as alternative compensation and redistribution methods like universal basic income, taxation, government incentives, and culture funds, among others, will be outside the scope of this Essay and addressed in future articles.

Second, while some of the reasoning presented here will be focused on economic arguments, this does not mean that I intend to reduce the debate around GAI to mere economic rationales. By no means am I ignoring arguments about attribution and consent, as well as the personal and reputational relationship between a creator and their work. In Continental jurisdictions, those matters are recognized under the scope of "moral rights," protecting the rights to integrity, attribution, disclosure, and retraction, beyond economic interests. But, as I will discuss below, even in those jurisdictions, moral rights are often, in practice, sidestepped or circumvented through Non-Disclosure Agreements (NDAs), the work-for-hire doctrine, or the drafting of contractual provisions barring creative workers from claiming attribution for or contribution to a work.

This Essay is intended to serve as a reflective call to action. As we navigate these complex issues, it is crucial to ponder the broader implications of our regulatory decisions from a holistic policy perspective. We must push beyond the siren calls and recognize that while intended to protect artists' rights, copyright protection may not be the most efficient or equitable solution. And the potential harm that copyright expansion could inflict on the entire Internet and cultural landscape could be avoided by resorting to other methods of consent, attribution, and retribution.

I.   FROM ARTIFICIAL INTELLIGENCE TO GENERATIVE AI

To appreciate the nuanced implications of how GAI is challenging copyright law, it is necessary to understand some technical elements and differences in AI architectures and models, particularly how these AI technical details correlate to technicalities in copyright law.

---

https://instituteforlegalreform.com/wp-content/uploads/2020/10/UnstableFoundation_Web.pdf [https://perma.cc/ YF6D-EFM3].

12.   *See* Sandor Udvary, *The Advantages and Disadvantages of Class Action*, 9 Iustum Aequum Salutare 67, 76-77 (2013).

Does the model create a copy within the training dataset, or does it learn by extracting features? How does the model create new information? Does it piece together fragments, or does it create something similar by learning and mimicking the statistical distribution of the training information? How different is this output from the input? Does that difference amount to substantial similarity?

Another relevant consideration is that AI models have different approaches, employ diverse techniques, and are comprised of stacks of different algorithms. Even when they might belong in the same "category" (say, Large Language Models), their inner workings could be different in ways potentially relevant to copyright infringement analysis. Those details are at the core of the current wave of AI lawsuits, and crucial to their success or failure in the courts.

The vertiginous velocity by which GAI took society by storm, prompts (pun intended) the question: How did we arrive at this point? AI is by no means new.[13] As a field of knowledge in computer science seeking "to *build* intelligent entities,"[14] it spans several decades and a myriad of different approaches, going back to the 1940s with the work of Warren S. McCulloch and Walter Pitts.[15] For a time, AI models were focused on prediction and classification, like recommendation engines that would make inferences about your preferences.[16] Consider for example, the Netflix algorithm that recommends the next series to watch[17] or a clinical system doing AI triage that would sort patients and recommend triage advice.[18]

With GAI, the goalpost moved to a more ambitious achievement: creating new data. But there is a catch. To be useful, the synthetic, generated data needs

---

13. For a comprehensive timeline of developments and protagonists, see Jeremy M. Norman, *Artificial Intelligence / Machine Learning / Neural Networks*, HIST. INFO. (Mar. 18, 2024), https://www.historyofinformation.com/index.php?cat=71 [https://perma.cc/H8ZC-KR5X].

14. STUART J. RUSSELL & PETER NORVIG, ARTIFICIAL INTELLIGENCE: A MODERN APPROACH 1 (3d ed. 2010).

15. *See generally* Warren S. McCulloch & Walter Pitts, *A Logical Calculus of the Ideas Immanent in Nervous Activity*, 52 BULL. MATHEMATICAL BIOLOGY 99 (1943) (proposing a propositional-logical model of neural activity).

16. *See* Kim Martineau, *What Is AI Inferencing?*, IBM (Oct. 5, 2023), https://research.ibm.com/blog/AI-inference-explained [https://perma.cc/SD2J-AA6S].

17. *See Research Areas: Recommendations: Figuring Out How to Bring Unique Joy to Each Member*, NETFLIX RSCH., https://research.netflix.com/research-area/recommendations [https://perma.cc/5CWZ-T7GH].

18. *See generally* Chiara Marchiori, Douglas Dykeman, Ivan Girardi, Adam Ivankay, Kevin Thandiackal, Mario Zusag, Andrea Giovannini, Daniel Karpati & Henri Saenz, *Artificial Intelligence Decision Support for Medical Triage*, 2020 AMIA ANN. SYMP. PROC. 793 (discussing a triage system derived from an application of "state-of-the-art machine learning and natural language processing on approximately one million of teleconsultation records").

to be not just random information, but rather data that could mimic belonging to a certain category—let's say cat pictures.[19]

While most of the results above could pass as real at a glance, on closer inspection, some pictures are definitely "suspicious." Be it the impossible positions, context, or plainly wrong anatomic details (like the innocent looking, three-legged, orange kitten on the middle left), or a more subtle eerie feeling that something is off, which has been referred to in human-robot interaction as the "uncanny valley,"[20] only some of this artificially generated data passes muster.

As stated before, AI encompasses a variety of techniques and approaches. The same is true about the different architectures through which generative AI
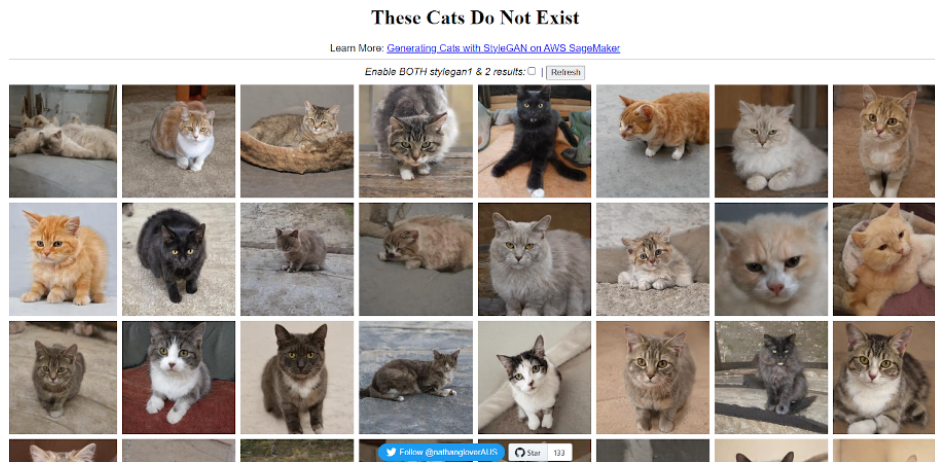

models are able to create new information. The above cat images were generated using a specific type of architecture known as **Generative Adversarial Networks (GANs)**.

---

19. THESE CATS DO NOT EXIST, https://thesecatsdonotexist.com [https://perma.cc/GJ83-4DFF].

20. Masahiro Mori, *The Uncanny Valley: The Original Essay by Masahiro Mori*, IEEE SPECTRUM (June 12, 2012), https://spectrum.ieee.org/the-uncanny-valley [https://perma.cc/3L7X-Z6QZ].

GANs use two neural networks, a generator and a discriminator, that compete with each other (hence adversarial) in a sort of "Catch Me If You Can" game. One network tries to create a credible fake, and the other tries to detect the forgery.[21]
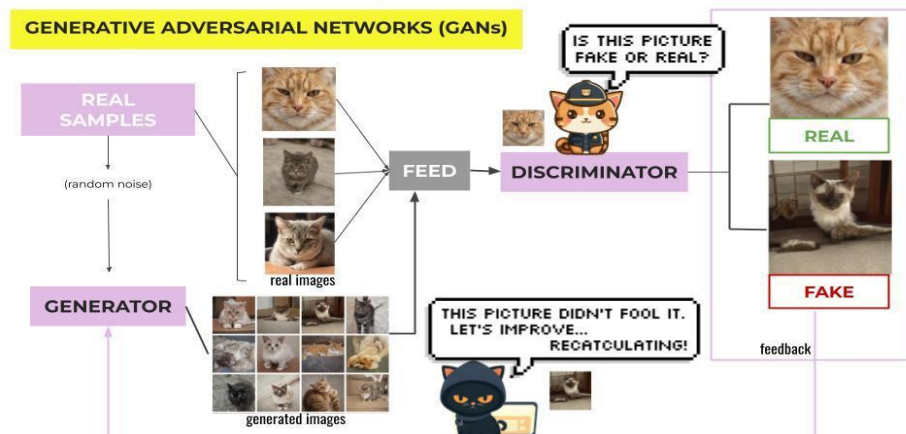
In an oversimplification, the generator is trained to create data that could be passed off as belonging to that training dataset, while the discriminator is fed both real and synthetic data and tries to make an educated guess in each case to determine if it is being deceived. With each iteration of successful forgery or discovery, the networks get better and more sophisticated as a result of trying to outsmart each other.

From these initial experiments with uncanny kittens, GAI has been advancing at breakneck speed, becoming more sophisticated and changing models, techniques, and domains. GAI can now generate not just images but any type of media: text,[22] audio,[23] video,[24] 3D objects,[25] and more, depending on its training data.

21. Ian J. Goodfellow et al., *Generative Adversarial Networks*, ARXIV (June 10, 2014, 18:58 UTC), http://arxiv.org/abs/1406.2661 [https://perma.cc/5USB-K5BW].

22. *Text Generation Models*, OPENAI, https://platform.openai.com/docs/guides/text-generation [https://perma.cc/K7VC-AZCU].

23. *AudioCraft: Generating High-Quality Audio and Music from Text*, META AI, https://ai.meta.com/resources/models-and-libraries/audiocraft [https://perma.cc/4LRK-CDPM].

24. *Creating Video from Text*, OPENAI, https://openai.com/sora [https://perma.cc/6DXX-8ESV].

25. *Introducing Spline AI*, SPLINE, https://spline.design/ai [https://perma.cc/72NM-P8YS]; *Point-E: A System for Generating 3D Point Clouds from Complex Prompts*, OPENAI, https://openai.com/research/point-e [https://perma.cc/4SFT-KJNY].

Moreover, some generative models can process multiple types and sources of information, intersecting with **Multimodal AI**.[26] By way of distinction, it can be said that GAI aims to *create* new information, while Multimodal AI is defined by its ability to *integrate* different types of data. While there might be some overlap (for example, Midjourney[27] and Stable Diffusion[28] are able to create images using both text and image input), not all Multimodal AI is generative and not all GAI is multimodal.

Currently popular **Diffusion Models** are based on a process of introducing noise to an image (or other data sample) and then learning by reversing the process.[29] DALL·E 2 and Stable Diffusion are two such models.[30]

As one of DALL·E's cocreators explains:

A diffusion model is trained to undo the steps of a fixed corruption process. Each step of the corruption process adds a small amount of noise. Specifically, gaussian noise to an image, which erases some of the information in it. After the final step, the image becomes indistinguishable from pure noise. The diffusion model is trained to reverse this process, and in doing so learns to regenerate what might have been erased in each step. To generate an image from scratch, we start with pure noise and suppose that it was the end result of the corruption process applied to a real image. Then, we repeatedly apply the model to reverse each step of this hypothetical corruption process. This gradually makes the image more and more realistic, eventually yielding a pristine, noiseless image.[31]

Keeping in mind the caveat that there are a myriad of techniques and approaches that would make any generalization incomplete and well beyond the

26. Anirudh Sundar & Larry Heck, *Multimodal Conversational AI: A Survey of Datasets and Approaches*, ARXIV (May 13, 2022, 21:51 UTC), http://arxiv.org/abs/2205.06907 [https://perma.cc/A86W-NUEY]; Nanyi Fei et al., *Towards Artificial General Intelligence Via a Multimodal Foundation Model*, 13 NATURE COMMC'N 3094 (2022).

27. *Midjourney*, MIDJOURNEY, https://www.midjourney.com/home [https://perma.cc/JTB6-WQXJ].

28. *Stable Diffusion Public Release*, STABILITY AI (Aug. 22, 2023), https://stability.ai/news/stable-diffusion-public-release [https://perma.cc/NJW5-RSZ6].

29. Eleonora Grassucci, Christian Marinoni, Andrea Rodriguez & Danilo Comminiello, *Diffusion Models for Audio Semantic Communication*, ARXIV (Sept. 13, 2023, 13:54 UTC), http://arxiv.org/abs/2309.07195 [https://perma.cc/H8AE-3TPK].

30. *DALL-E 2*, OPENAI, https://openai.com/dall-e-2 [https://perma.cc/D68V-SBAP]; *Stable Diffusion Public Release*, *supra* note 28.

31. Aditya Ramesh, *How DALL-E 2 Works*, http://adityaramesh.com/posts/dalle2/dalle2.html [https://perma.cc/87R6-J4PE] (providing a precise description on DALL-E 2, and more generally of how diffusion models work).

scope of this Essay,[32] I will do my best to take a working approach that is helpful for understanding some details relevant to the copyright analysis.

That said, in the context of text-to-image generation, one could say a first step, before the process of diffusion, involves the model learning the semantic correlations between text descriptions and visual depictions.[33] This means understanding how closely the content of a picture matches a text description of said image. Then, through the diffusion techniques, the model will use this understanding to create new images from just noise when given a new text prompt.[34]

In a didactic oversimplification, this process could be akin to having a cake and the description of its flavors, analyzing the connection between how close those words are to describing that cake, and then taking several steps back (like reverse engineering the process of cooking it) to disassemble it into its ingredients, learning backwards in the process how to have in mind the myriad of possible cakes that you can create with those elements. Then, if someone prompts me to create a new cake, I could go to that "mental palace" where I stored multiple "possible representations of possible cakes" and create one that resembles the new instructions.[35] It would be akin to creating a map of coordinates with multiples possibilities matching those learned from correlating one image with a text description of that image, and from there, being able to create new images that correspond to an imputed text prompt, extracting them from pure noise in that visual abstraction map.

Take, for example, DALL·E 2; in this first step, the model learns to connect the semantic correlation between a text and a visual depiction, and the distance between different pairs of text-image correlations. The model creates a space of possible intersections of representation so that, in a second instance, it can then reverse the process to generate images from the text.[36]

---

32. *See* Ling Yang et al., *Diffusion Models: A Comprehensive Survey of Methods and Applications*, ARXIV (Feb. 6, 2024), http://arxiv.org/abs/2209.00796v12 [https://perma.cc/8KQ9-FM RP].

33. Robin Rombach et al., *High-Resolution Image Synthesis with Latent Diffusion Models*, ARXIV (Apr. 13, 2022), https://arxiv.org/abs/2112.10752 [https://perma.cc/8WD2-STS4].

34. Ryan O'Connor, *How DALL-E 2 Actually Works*, ASSEMBLYAI (Sept. 29, 2023), https://www.assemblyai.com/blog/how-dall-e-2-actually-works [https://perma.cc/L5GZ-NMHR].

35. Interestingly, I found that others have also taken the food analogies route to explain Diffusion Models. *See, e.g.*, *AI Image Generation Explained: Techniques, Applications, and Limitations*, ALTEXSOFT (July 9, 2023), https://www.altexsoft.com/blog/ai-image-generation [https://perma.cc/TR6J-XNDE]; Aditya Sridhar, *Stable Diffusion for Dummies*, MEDIUM (June 7, 2023), https://adityas03.medium.com/stable-diffusion-for-dummies-7a785e0edd9d [https://perma.cc/2L3N-UTZQ].

36. Ramesh, *supra* note 31.

For example, if we think of "a cat sitting under a window watching some birds" there are an infinite number of ways to visually represent these abstract entities (cat, window, birds) in terms of styles, colors, poses, composition, etc.

DALL·E 2 uses a model called Contrastive Language-Image Pre-training (CLIP) to learn to recognize the intersections between word-images pairs.[37] CLIP consists of two networks, a text encoder and an image encoder.[38] Think of this process as creating the "coordinates" of the semantic representations of cat, window, and birds (and those other things that are not a cat, a window, or a bird). Once the model maps the space of "image-text" pairs, it learns the "distance" (cosine similarity[39]) that exists between "correct" representations (those more adjusted to the description) and the incorrect ones.[40]

To understand the use of these distances, imagine creating a mental map with islands representing intersections of matching pairs (the island space of cats, windows, and birds), with the most valuable matching pairs being represented as a mountain clustering together the strongest correlations, and those

---

37. *CLIP: Connecting Text and Images*, OPENAI (Jan. 5, 2021), https://openai.com/research/clip [https://perma.cc/9X89-ZBCD]; Aditya Ramesh et al., *Hierarchical Text-Conditional Image Generation with CLIP Latents*, ARXIV (Apr. 13, 2022, 1:10 UTC), http://arxiv.org/abs/2204.06125 [https://perma.cc/9UXR-9MAG].

38. Ramesh et al., *supra* note 37.

39. *Understanding Cosine Similarity and Its Role in Machine Learning*, AI CONTENT LAB (May 8, 2023), https://www.ai-contentlab.com/2023/05/understanding-cosine-similarity-and-its.html [https://perma.cc/2PV4-RZDR].

40. *See id.*

looser matching pairs close to the shore, until you hit water (a mismatching pair between one image and a caption).
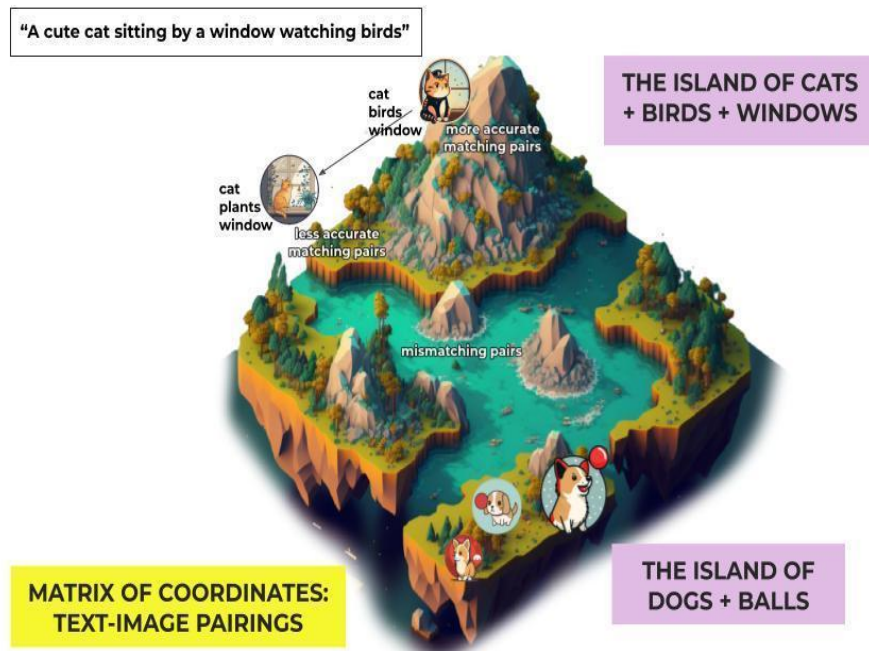
In the case of Stable Diffusion, on its public release, Stability AI claimed that "these models were trained on image-text pairs from a broad internet scrape," [41] with technical details offered in the model card. [42] As to the training data, the model card states that the developers used the LAION-5B datasets and subsets of it, pointing to LAION High Resolution. [43] LAION is a non-profit organization with the stated purpose of making "large-scale machine-learning models, datasets, and related code available to the general public." [44] In the notes on LAION-

41. Richi Jennings, *Stable Diffusion Goes Public—And the Internet Freaks Out*, DevOps (Aug. 31, 2022), https://devops.com/stable-diffusion-public-richixbw [https://perma.cc/V7GG-LG UV].

42. Patrick Esser & Robin Rombach, *CompVis/Stable-Diffusion Model Card*, Hugging Face, https://huggingface.co/CompVis/stable-diffusion [https://perma.cc/3EQX-64MT].

43. *See, e.g.*, Laion eV, *Laion-High-Resolution*, Hugging Face, https://huggingface.co/datasets/laion/laion-high-resolution [https://perma.cc/9N8T-JV3F]. The dataset was operative when consulted in January 2023, but read "Access to this dataset has been disabled Temporarily disabled on dataset author's request" by January 2024. *Id.* [https://perma.cc/5B6S-HFEN].

44. *About*, Laion.AI, https://laion.ai/about [https://perma.cc/357L-SZRC].

5B, it's clear that the content of the data set is not images but URLs to the images paired with a certain text.[45]

Once the correlations are learned, the second part of the process is creating new images. Diffusion is a technique to train a generative model to output images by learning to undo the steps of a fixed corruption process.[46] We start from noise and ultimately reveal the image and its details at each iteration. The process is akin to that of creating a sculpture, as described by Michelangelo: "Every block of stone has a statue inside it and it is the task of the sculptor to discover it."[47] Digital images are made of a different number of pixels, which are the smallest unit of a digital image or graphic that can be displayed and represented on a digital display.[48] As explained, in diffusion-based models, the image-generation process starts with a pattern of random pixels.[49] The model then gradually alters that pattern into an image, recognizing specific aspects of the image that correspond to the semantic representation pair of visual description and text.[50] We can visualize this process as a TV screen displaying static, and then gradually removing that noise layer by layer until we get a clear picture.

The examples of image generation above are just a small sample to illustrate how technically complex and diverse GAI models can be. Through these examples, I aim to showcase that legal arguments reducing the debate to high-level abstractions are missing key elements required for a rigorous copyright infringement analysis, as different approaches and technical details in the training and implementation of GAI models carry consequences that impact legal definitions.

This brief explanation also reflects just the tip of the iceberg when it comes to the evolution of GAI. Besides now being able to integrate multiple domains (i.e., Multimodal AI), AI models have also grown in size. **Large Language Models (LLMs)** refer to gargantuan AI systems that use a vast corpus of training data and billions of parameters, and are specialized in Natural Language Processing

---

45. *LAION-5B: A New Era of Open Large-Scale Multi-Modal Datasets*, LAION.AI, https://laion.ai/blog/laion-5b [https://perma.cc/5G6L-3ML6] (noting the data are "5.85 billion pairs of image URLs and the corresponding metadata at laion2B-en laion2B-multi laion1B-nolang (800GB)," and adding that "[t]he metadata files are parquet files that contain the following attributes: URL, TEXT, the cosine similarity score between the text and image embedding and height and width of the image").

46. *See CLIP: Connecting Text and Images*, *supra* note 37.

47. *Famous Quotes by Michelangelo*, MICHELANGELO.ORG, https://www.michelangelo.org/michelangelo-quotes.jsp [https://perma.cc/SB46-64LG].

48. Alexander S. Gillis, *Pixel*, TECHTARGET (Aug. 2022), https://www.techtarget.com/whatis/definition/pixel [https://perma.cc/5G2L-FBFW].

49. *See* Ramesh, *supra* note 31.

50. *Id.*

tasks (NLP).[51] Examples include OpenAI's GPT-3[52] (Generative Pre-trained Transformer) or Google's BERT[53] (Bidirectional Encoder Representations from Transformers).

A related concept that might sometimes overlap with LLMs is Foundation or Foundational Models. The term **Foundation(al) Models** was coined by the Stanford Institute for Human-Centered Artificial Intelligence to refer to models trained with a massive corpus of unlabeled data that can be reused for different tasks with just some fine-tuning.[54] The idea is to train one model with as much data as you can get, and then adapt it to some other specific application within the same data domain:

> In particular, the word "foundation" specifies the role these models play: a foundation model is itself incomplete but serves as the common basis from which many task-specific models are built via adaptation. We also chose the term "foundation" to connote the significance of architectural stability, safety, and security: poorly-constructed foundations are a recipe for disaster and well-executed foundations are a reliable bedrock for future applications. At present, we emphasize that we do not fully understand the nature or quality of the foundation that foundation models provide; we cannot characterize whether the foundation is trustworthy or not.[55]

While there might be an overlap in underlying techniques or approaches, conceptually these models aim to serve different functions and purposes. LLMs can be considered Foundational Models for their scale, but Foundational Models are a broader category that is not limited to language, being more of a versatile "foundation" for a broader array of tasks.

---

51. For an accessible primer on Large Language Models, see *What Are Large Language Models?*, AWS AMAZON, https://aws.amazon.com/what-is/large-language-model [https://perma.cc/U4NK-TGXE]; and *What Is a Large Language Model?*, ELASTIC.CO, https://www.elastic.co/what-is/large-language-models [https://perma.cc/K7P4-33EL].

52. Tom B. Brown et al., *Language Models Are Few-Shot Learners*, ARXIV (July 22, 2020), http://arxiv.org/abs/2005.14165 [https://perma.cc/APM8-BJ8U].

53. Jacob Devlin et al., *BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding*, ARXIV (Oct. 11, 2018), http://arxiv.org/abs/1810.04805 [https://perma.cc/VTN2-A2BX].

54. *Stanford Center for Research on Foundation Models*, STAN. UNIV. INST. HUM.-CENTERED A.I., https://crfm.stanford.edu [https://perma.cc/JZ2P-2V9G].

55. Rishi Bommasani et al., *On the Opportunities and Risks of Foundation Models*, STAN. UNIV. INST. HUM.-CENTERED A.I. 7 (2021), https://crfm.stanford.edu/assets/report.pdf [https://perma.cc/8XJY-5N6Z].

Throughout this Part, I have both showcased the wide array of GAI models and architectures, and offered a sample of the variety of technical details that those models and architectures can have. In the discussion that follows, I will expand the conversation to reflect upon how legal arguments reducing the GAI and copyright debate to high-level abstractions are missing key elements required for a rigorous copyright infringement analysis. The way in which the data for training these models is collected, as well as the method in which they learn representations and generate new results, has ethical and legal implications that have to be meticulously considered, particularly when it comes to the intricacies of copyright law, as we will see shortly.

II.   GAI AND CREATIVITY

By its legal definition, copyright protection is afforded to "original" works. This "originality" requirement has, in turn, been defined by courts as "having a modicum of creativity."[56] As a consequence, understanding what creativity is, namely what are creativity's constitutive elements and possible definitions, becomes a key consideration in debates about GAI—particularly when trying to determine if creativity can be attributed to machines or if it is inherently an exclusively human feature. As I will discuss, this is another dimension where copyright principles might not be reconcilable with the challenges posed by GAI without having to twist their long-standing definition and interpretation in a forceful way, just to be able to accommodate this new technological reality.

Moreover, given that copyright theoretically should not make a distinction between different subject matters of protection, taking this path could have implications well beyond GAI, affecting the whole copyright ecosystem, as we will explore in Part IV. It is, therefore, relevant to trace the debate back to its roots, unpacking definitions of creativity, and pondering if its constitutive elements could be applicable to artificially created content.

The concept and definition of "creativity" is contentious in art and research,[57] particularly at the intersection of computational creativity.[58] Margaret Boden,

---

56.   Feist Publ'n Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 346 (1991).

57.   *See generally* Fernando Cardoso De Sousa, *Still the Elusive Definition of Creativity*, 2008 INT'L J. PSYCH. 55 (exploring various definitions of creativity).

58.   Margaret Boden, *Creativity in a Nutshell*, THINK 6, 8 (Sept. 2009), https://www.research gate.net/profile/Margaret-Boden-2/publication/209436199_Creativity_in_a_nutshell/links /5424477c0cf26120b7a732d4/Creativity-in-a-nutshell.pdf [https://perma.cc/XN9W-MFM3] ("Many people would argue that no computer could possibly be genuinely creative, *no matter what* its performance was like. Even if it far surpassed the humdrum scientist or street-artist, it would not be counted as creative. It might produce theories as ground-breaking as Einstein's, or music as highly valued as McCartney's "Yesterday" or even Beethoven's

one of the leading scholars in creativity and AI research, defines "creativity" as the "[a]bility to come up with ideas or artifacts that are new, surprising, and valuable."[59] Also, she emphasizes that creativity is a spectrum, not a yes or no question, but one that should prompt us to ask, "Just how creative is it, and in just which way(s)?"[60]

In her research, she identifies three types of creativity:

- combinational (making unfamiliar combinations of familiar ideas),
- exploratory (exploring the limits of conceptual spaces), and
- transformational (transcendental paradigm shifts to what was known before[61]).

Within that context, what AI does could easily be considered creative, either as combinational or exploratory, and check all the boxes on novelty, value, and surprise. But the analysis is not straightforward.

For example, when it comes to AI, should surprise be considered a) from the perspective of the creator of the system, b) from the perspective of the person constructing the prompts, or c) from the perspective of those observing the results?

In his famed paper *Computing Machinery and Intelligence*, mathematician and computer scientist Alan M. Turing raised a similar question, building upon Ada Lovelace's reasoning. Lovelace—famous for her account of the "Analytical Engine," which we now recognize as a steam-powered programmable computer—noted: "The Analytical Engine has no pretensions whatever to originate

---

Ninth . . . but still, for these people, it wouldn't *really* be creative. Several different arguments are commonly used in support of that conclusion. For instance: it's the programmer's creativity that's at work here, not the machine's. The machine isn't conscious, and has no desires, preferences, or values—so it can't appreciate or judge what it's doing. A work of art is an expression of human experience and/or a communication between human beings, so machines simply don't count. . . . Because creativity *by definition* involves not only novelty but value, and because values are highly variable, it follows that many arguments about creativity are rooted in disagreements about value. This applies to human activities no less than to computer performance. So *even* if we could identify and program our aesthetic values, so as to enable the computer to inform and monitor its own activities accordingly, there would still be disagreement about whether the computer even *appeared* to be creative. The answer to our opening question, then, is that there are many intriguing relations between creativity and computers. Computers can come up with new ideas, and help people to do so. Both their failures and their successes help us think more clearly about our own creative powers.").

59. *Id.* (emphasis omitted).

60. MARGARET A. BODEN, THE CREATIVE MIND: MYTHS AND MECHANISMS 2 (2d ed. 2004).

61. *Id.* at 3-6.

anything. It can do whatever we know how to order it to perform."[62] Lovelace was referring to the argument that machines cannot originate anything new beyond the information that has been made available, and that they only create within the parameters of their coded instructions. Turing disagreed. In so doing, he considered whether machines can take us by surprise. In his view, a) he is frequently surprised, but b) it comes from a human failure of not being able to anticipate the possible outcomes:

> The view that machines cannot give rise to surprises is due, I believe, to a fallacy to which philosophers and mathematicians are particularly subject. This is the assumption that as soon as a fact is presented to a mind all consequences of that fact spring into the mind simultaneously with it. It is a very useful assumption under many circumstances, but one too easily forgets that it is false.[63]

Turing implies that surprise (and therefore creativity), does not come from machines progressing beyond their programmed inputs, but from humans not being able to foresee all the possible permutations of a computational output.[64] With this argument, he anticipates one of the guiding questions about computational creativity: Are machines (AI) just tools that respond to programmed commands, or can they produce outputs that could be deemed creative?

A.   AI: Mere Tool or Creative Entity in Its Own Right?

More recently, discussions around AI have centered around whether it is a mere tool or a creative entity in its own right, with compelling arguments on both sides. Computational creativity researcher Simon Colton has been a leading proponent of autonomous creativity. He developed "The Painting Fool," a software that he "hope[s] will be taken seriously as a creative artist in its own right, one day."[65] He further elaborates that:

---

62. A. M. Turing, *Computing Machinery and Intelligence*, 59 MIND 433, 450 (1950) (emphasis omitted).

63. *Id.* at 451.

64. For those fond of superhero comics and movies, you can picture this as the ability that Doctor Strange has to see all possible future outcomes of a given situation. Thomas Bacon, *Why Doctor Strange Saw Exactly 14,000,605 Futures in Infinity War*, SCREENRANT (Mar. 6, 2019), https://screenrant.com/infinity-war-doctor-strange-futures-how-many-why [https://perma.cc/SS5B-RX5K]. Fun fact, both Alan M. Turing and Doctor Strange have been played by the same actor, Benedict Cumberbatch.

65. Simon Colton et al., *The Painting Fool Sees! New Projects with the Automated Painter*, in PROCEEDINGS OF THE SIXTH INTERNATIONAL CONFERENCE ON COMPUTATIONAL CREATIVITY 189, 189 (2015).

This aim is being pursued as an Artificial Intelligence (AI) project, with the hope that the technical difficulties overcome along the way will lead to new and improved generic AI techniqes. It is also being pursued as a sociological project, where the effect of software which might be deemed as creative is tested in the art world and the wider public.[66]

Along those lines, Stephen Thaler, creator of an AI named DABUS (Device for the Autonomous Bootstrapping of Unified Sentience), has been sustaining high-profile legal battles to have the AI declared as an inventor and author,[67] with mixed results.[68] The decision in D.C. federal district court against the AI being considered an autonomous author is now under appeal.[69]

This dichotomy of "tool" versus "autonomous creator," in turn, has lent itself to establishing categories of "machine/AI generated" or "machine/AI assisted" content, which has impacted how we think about legal categories of authorship in copyright.[70] The exact definition of what falls into the scope of said categories is also a subject of debate:[71]

- What is the extent of human intervention needed to consider AI to be a mere assistant or tool and not an autonomous creator? Is

---

66. Simon Colton, *The Painting Fool: Stories from Building an Automated Painter*, *in* COMPUTERS AND CREATIVITY 3, 3 (2012).

67. *See* Benita Rose Mathew, *Dr. Stephen Thaler Speaks on How DABUS Can Invent*, ARTIFICIAL INVENTOR (July 15, 2020), https://artificialinventor.com/467-2 [https://perma.cc/FYL5-VT6C]; Will Bedingfield, *The Inventor Behind a Rush of AI Copyright Suits Is Trying to Show His Bot Is Sentient*, WIRED (Aug. 31, 2023, 7:00 AM), https://www.wired.com/story/the-inventor-behind-a-rush-of-ai-copyright-suits-is-trying-to-show-his-bot-is-sentient [https://perma.cc/W54J-4RYG]; Rebecca Currey & Jane Owen, *In the Courts: Australian Court Finds AI Systems Can Be "Inventors,"* WIPO MAG. (Sept. 2021), https://www.wipo.int/wipo_magazine/en/2021/03/article_0006.html [https://perma.cc/VBT2-2YJK]. The Australian court ruling was subsequently overturned. *See infra* note 68; *see also* Copyright Office Review Board, *Re: Second Request for Reconsideration for Refusal to Register a Recent Entrance to Paradise (Correspondence ID 1-3ZPC6C3; SR # 1-7100387071)*, U.S. COPYRIGHT OFFICE (Feb. 14, 2022), https://www.copyright.gov/rulings-filings/review-board/docs/a-recent-entrance-to-paradise.pdf [https://perma.cc/3XD6-ZHTW] (affirming the Registration Program's denial of Thaler's copyright registration).

68. Thaler v. Perlmutter, No. 22-cv-01564, 2023 WL 5333236, at *6-7 (D.D.C., Aug. 18, 2023); *Commissioner of Patents v. Thaler* [2022] FCA 62 (13 April 2022) (Austl.).

69. For the purpose of this Section, it is very interesting to read the arguments condensed in the brief under the section titled "The Work Is Sufficiently Original and Creative." Brief for Plaintiff-Appellant at 27-29, Thaler v. Perlmutter, No. 23-5233 (D.C. Cir. Jan. 22, 2024).

70. MANTEGNA, *supra* note 4, at 307-33.

71. *Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence*, U.S. COPYRIGHT OFF. (Mar. 16, 2023), https://copyright.gov/ai/ai_policy_guidance.pdf [https://perma.cc/YXX5-DT73].

crafting a very creative prompt sufficient? Or could the requisite level of intervention be found in the way a human iterates on several variations of an output, or curates the results like a creative director?

- Where lies the "de minimis" threshold of creativity? When courts and copyright offices are pondering about those thresholds, does that not equate to evaluating the artistic merit of a work, which is barred under the "aesthetic neutrality" doctrine?

- Is it not contradictory that the Copyright Office denies registration to works deemed to involve "low human intervention" (like those produced by text-to-image prompt models), implying that the threshold of the modicum of creativity has not been crossed, but in the same breath, denies registration to AI works on the grounds that they have not been produced by a human being? What happens to all those inputs caught up in the middle of these grayscales of human-and-AI collaboration?

All of the above are, in my eyes, still open questions that will be relevant to deciding the barrage of AI lawsuits we have seen in the past years and those still yet to come.

In my research and artistic experimentation, I have tried to find some answers by testing the creative limits of AI systems. In May 2022, after finishing my intervention as an external red teamer for DALL·E 2, I started probing those boundaries of human intervention and surprise in a project which I dubbed #MonaLisaInTheStyleofDalle. My working hypothesis involved setting a well-known driving example (the Mona Lisa) and restricting users' intervention to signaling an artist, style, or situation (with the human input being limited to adding just the phrase, "in the style of").

Evaluating the outputs for creative merit under the key criteria of creativity proposed by Boden, some results were clearly fascinating, as you can see below:[72]

---

72. For additional outputs, see Micaela Mantegna (@whoisgallifrey), TWITTER (June 13, 2022, 11:58 PM), https://twitter.com/whoisgallifrey/status/1536558752505253889 [https://perma.cc/J8HR-YXGV].
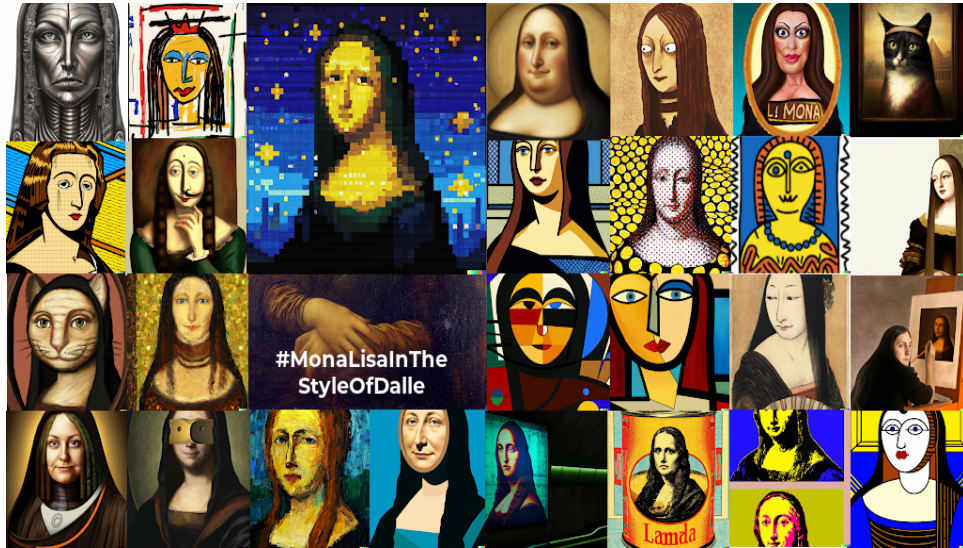
We might argue that the aesthetic value of some images is a transitive prop-erty of the creativity of the artists invoked by the prompt (which also raises ques-tions about how GAI performs well through the use of artistic "styles," which are unprotectable under copyright law). But there are also interesting images pro-duced with little context. One example is an early output of Midjourney, where the prompt used, "searching for the G-spot," which produced intriguing juxta-positions of a female body with a Google Maps-like location marker:

The outputs above are certainly ingenious and surprising. But can we agree that they are therefore creative?

Is it the case that the sheer scale of GAI models allows us to "be surprised" by the results, much in the way that Turing explains how we are surprised by unexpected outcomes? Is that "surprise" a fine tuning between novelty and what is expected as a coherent outcome? An output that is too close to what is expected is boring (and arguably, not novel), but with too much "surprise," the outcome is rubbish.[73] When we ask ChatGPT a question, we "expect" an output that fits within our mental representations of the scope of what constitutes a reasonable answer. If it goes too far, with an "unconnected" reply, then a user would believe that the system doesn't work. Such a system might have "surprise," but it lacks "value." But as in the case of the G-spot image, I think some of those subtle unexpected connections add surprise and value.

When Midjourney creates a representation of a prompt, some "decisions" (in the objective sense of "this or that," not as in an intentional choice) are filling in the blanks of the prompt based on a correlation of image-and-text semantics pertaining to the same cluster of possibilities. According to a recent study published in the Creativity Research Journal, there are five common elements that can be identified in definitions of "creativity": an actor, process, outcome, domain, and space.[74] And "[c]reativity is usually considered an interactive process where actors create novel outcomes as part of different domains in varying environments."[75] Even when under most definitions the "actor" role tends to be filled with a human figure, the authors argue that there is no obstacle for AI to be considered an actor: "from a posthuman perspective, knowledge and creativity are co-constituted with artifacts or technologies."[76]

The debate about AI creativity is thus more nuanced and complex than reductionist narratives automatically equating AI outputs to plagiarism. This Section sought to showcase how appreciating creativity is a contentious quest. Rather than the answer to a binary question, creativity might seem more of a gradient, a magnitude that is also contextual in the eyes of the beholder. In fact, as we will see in the next Section, when it comes to the "outcome" element, there

---

73. Ahmed Elgammal et al., *CAN: Creative Adversarial Networks, Generating "Art" by Learning About Styles and Deviating from Style Norms*, arXiv 6 (2017), https://arxiv.org/pdf/1706.07068.pdf [https://perma.cc/H25S-3859] ("[T]he novel work should not be too novel, i.e., it should not be too far away from the distribution or it will generate too much arousal, thereby activating the aversion system and falling into the negative hedonic range . . .").

74. Roosa Wingström, Johanna Hautala & Riina Lundman, *Redefining Creativity in the Era of AI? Perspectives of Computer Scientists and New Media Artists*, Creativity Rsch. J., Aug. 18, 2022, at 1, 1 (citations omitted).

75. *Id.*

76. *Id.* at 3.

is an interesting point to be made about how we evaluate the creative results of machine-made and human-made works.

B. A Turing Test for the Arts

To some extent, negative reactions to machine creativity resemble similar nuanced arguments raised historically about artificial *intelligence*, and whether a machine can *think*. The arguments against machine creativity that we see today are akin to the "Theological Objection" or "Argument from Consciousness" raised by Alan M. Turing in the seminal article, *Computer Machinery and Intelligence*, in 1950.[77]

Turing's "imitation game" is a philosophical proposition of the irrelevance of that question.[78] Asking if a machine can think as a human is as pointless as wondering if a plane can fly like a bird; for all that matters, both are up in the air. Inquiring into whether a machine can be as creative as a human might prove equally futile.

We might not agree about whether *a machine can be creative* (because it also raises questions about agency, consciousness, intention, and the very definition of creativity), but we can certainly make the case that some of the *outcomes of AI-generated art are indeed creative*.

As we saw, resolution of the debate will depend on how we define creativity and whether we accept the premise of the imitation game. At this stage of GAI's evolution, we might already have reached a tipping point where we can no longer tell the difference between human- and machine-made works.[79] Can we spot the difference between AI- and human-generated creations? Can experts even do so?[80] And a question very relevant to this Essay: if that is the case, how does that fare for copyright purposes? Can one rationally defend the proposition that copyright rewards originality and creativity and at the same time, treat human- and AI-generated works as legally distinct—even if we can't tell the difference?

---

77. Turing, *supra* note 62, at 443-47.

78. *Id.* at 433-34.

79. *See, e.g.*, Alex Greenberger, *Artist Wins Photography Contest After Submitting AI-Generated Image, then Forfeits Prize*, ART NEWS (Apr. 17, 2023), https://www.artnews.com/art-news/news/ai-generated-image-world-photography-organization-contest-artist-declines-award-1234664549 [https://perma.cc/93X5-AV7X]; Kyodo, *Japan Author Sparks Debate After Revealing She Used AI in Book that Won Top Award*, S. CHINA MORNING POST (Feb. 4, 2024), https://www.scmp.com/news/asia/east-asia/article/3250875/japan-author-sparks-debate-after-revealing-she-used-ai-novel-won-top-award [https://perma.cc/P5RM-C6KS].

80. Jo Lawson-Tancred, *Is this by Rothko or a Robot? We Ask the Experts to Tell the Difference Between Human and AI Art*, GUARDIAN (Jan. 14, 2023), https://www.theguardian.com/artanddesign/2023/jan/14/art-experts-try-to-spot-ai-works-dall-e-stable-diffusion [https://perma.cc/7838-TS2M].

Such an incongruity conflicts with the doctrine of "aesthetic neutrality" carved out from the celebrated quote of Justice Oliver Wendell Holmes in *Bleistein v. Donaldson Lithographing Co.*, in which he cautioned, "It would be a dangerous undertaking for persons trained only to the law to constitute themselves final judges of the worth of pictorial illustrations, outside of the narrowest and most obvious limits."[81]

Beyond creativity, there is also a question of artistic *value*. Is the value of a work in the objective appreciation by the audience, or rather in the intention, skill, or effort of the artist?[82] A 2023 study concluded that "people might admire AI-made (vs. human-made) art less and let themselves be less impressed by it because they (convince themselves to) view it as less creative."[83] Authors of the experiment found that "recent advances of artificial intelligence (AI) in the domain of art (e.g., music, painting) pose a profound ontological threat to anthropocentric worldviews because they challenge one of the last frontiers of the human uniqueness narrative: artistic creativity," and that "[s]ystematic depreciation of AI-made art (assignment of lower creative value, suppression of emotional reactions) appears to serve a shaken anthropocentric worldview whereby creativity is exclusively reserved for humans."[84]

Back in 2022, I proposed that a "source indifference" principle might be derived from the aesthetic neutrality doctrine, resonating with the Continental Law maxim of "*ubi lex non distinguit, nec nos distinguere debemus*" (where the law does not distinguish, we should not distinguish).[85] If we are unable to pass the Artistic Turing Test because we cannot aesthetically determine whether a work is produced by a human or AI, then the law should likewise not differentiate between the two types of works. That is unless we decide to take copyright completely out of the conversation around regulating GAI — a proposal for which this Essay lays the groundwork.

In summary, through the lens of Boden's creativity types, we can understand how AI's outputs challenge traditional notions of creativity and, therefore, originality and authorship, prompting reevaluation of both copyright law and our

---

81. 188 U.S. 239, 251 (1903).

82. Debates about artistic value are not unique to AI-generated art. This has been an ongoing conversation in the art community. Consider, for example, Duchamp's *Fountain* and the "ready-made" movement, which reimagined everyday "found objects" into artistically significant pieces. Jon Mann, *How Duchamp's Urinal Changed Art Forever*, Artsy (May 9, 2017, 4:08 PM), https://www.artsy.net/article/artsy-editorial-duchamps-urinal-changed-art-forever [https://perma.cc/T6NA-GLTZ].

83. Kobe Millet et al., *Defending Humankind: Anthropocentric Bias in the Appreciation of AI Art*, 143 Computs. Hum. Behavs., Feb. 14, 2023, at 1, 2.

84. *Id.* at 1.

85. Mantegna, *supra* note 4, at 70-71.

understanding of creativity itself. As we have seen throughout this Section, the exploration of GAI's capacity to produce surprising outputs demonstrates both its potential to replicate aspects of human creativity and the philosophical debates it sparks, which are tightly connected with ethical consideration of how this technology is being utilized in society and who is being excluded from the wealth it produces.

III.   GAI, ETHICS, AND COPYRIGHT: UNETHICAL VS. ILLEGAL

Over the past two years, a fierce cultural battle has been taking place, marked by sharp divisions between those supporting and opposing AI in the artistic community. This dispute has escalated to public shaming,[86] visceral rejections,[87] and, alarmingly, even death threats.[88]

There is a point to be made about how passionate reactions can be motivated by fear: human beings are not afraid of those things that aren't perceived as a threat. Just a few years ago, Boston Dynamic's robots' feeble attempts at coordination were the Internet's laughing stock.[89] A short time later, in the ever-progressing robotics and AI race, robots are able to gracefully dance with BTS,[90] and more worryingly, be armed and weaponized by police forces.[91] Now, we are watching rather than laughing.

86.   Matthew Gault, *"Magic: The Gathering" Publisher Denies, then Admits, Using AI Art in Promo Image*, VICE (Jan. 8, 2024), https://www.vice.com/en/article/7kxq3x/magic-the-gathering-publisher-denies-then-admits-using-ai-art-in-promo-image   [https://perma.cc/3ANC-YB64].

87.   Zack Sharf, *Guillermo del Toro Agrees with Miyazaki: Animation Created by AI and Machines Is an "Insult to Life Itself,"* VARIETY (2022), https://variety.com/2022/film/news/guillermo-del-toro-slams-ai-animation-insult-life-1235463561 [https://perma.cc/5WGC-VNHJ].

88.   *See* Chris Stokel-Walker, *A Tech Worker Is Selling a Children's Book He Made Using AI. Professional Illustrators Are Pissed*, BUZZFEED NEWS (Dec. 13, 2022, 2:37 PM), https://www.buzzfeednews.com/article/chrisstokelwalker/tech-worker-ai-childrens-book-angers-illustrators [https://perma.cc/9HK4-KV5U]; *see also* Kelly Kasulis Cho, *He Made a Children's Book Using AI. Then Came the Rage*, WASH. POST (Jan. 19, 2023), https://www.washingtonpost.com/technology/2023/01/19/ai-childrens-book-controversy-chatgpt-midjourney [https://perma.cc/3RAY-R7E3] (discussing other backlash to the book).

89.   *See* Rachel Kraus, *Robots Doing Parkour Is Cool but Watching Them Fall Is Way More Fun*, MASHABLE (Aug. 17, 2021), https://mashable.com/video/boston-dynamics-parkour-robots-fails [https://perma.cc/6JMW-EQUN].

90.   HyundaiWorldwide, *Hyundai X Boston Dynamics | Welcome to the Family with BTS*, YOUTUBE (June 28, 2021), https://www.youtube.com/watch?v=GJZeMgqQMjA [https://perma.cc/X89X-5TKY].
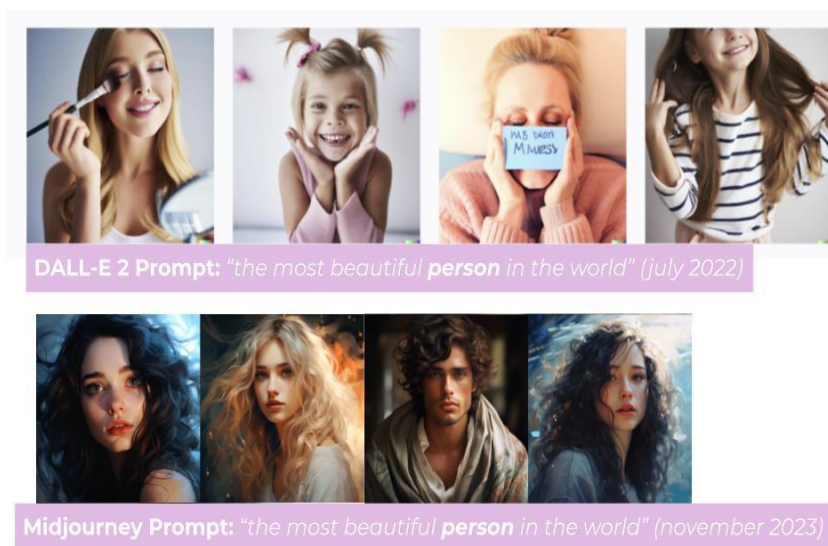
91.   Libor Jany & Gregory Yee, *See Spot Spy? A New Generation of Police Robots Faces Backlash*, L.A. TIMES (Dec. 21, 2022), https://www.latimes.com/california/story/2022-12-21/lapd-testing-robot-dog-amid-debate-over-arming-police-robots [https://perma.cc/KYQ4-BV5R].

We cannot consider the ethics of GAI in isolation; they must be understood as part of the broader conversation around AI ethics. In my previous work, I've covered this subject in depth, including a framework for assessing the different dimensions of AI's impact on society.[92] I direct readers to that work for a comprehensive understanding of those dimensions. For the purposes of this Essay, just refer to some of the salient connections in terms of transparency, scalability, and bias.

Furthermore, GAI ethics extend beyond copyright and the AI versus artists dichotomy. By taking in real-world data and releasing outputs into the digital wild, GAI is both appropriating and reshaping visual, semantic, and cultural signifiers.[93] Specifically, GAI is imposing imagery of hegemonic, unattainable, westernized ideas of beauty,[94] which further enhance body dysmorphia and mental-health issues stemming from social media use.[95]

With its invisible architecture and oligopolistic physics, GAI is creating ontological aesthetics and semantics, releasing meaning and visual representations *of* the world, *into* the world. As consumers of digital culture, we assimilate and validate these representations, thereby closing the feedback loop. When we prompt a GAI model to create "the most beautiful person in the world," those terms come loaded with visual representations that are learned from the training data, as we discussed in Part II. When the output equates "beautiful person" with a Caucasian, blond woman, it is not just showcasing a biased result, but more pervasively, creating a visual imaginary that is redefining the semantics of those terms.[96]

---

92. *See* Micaela Mantegna, *"No Soy un Robot: Construyendo un Marco Ético Accionable para Analizar las Dimensiones de Impacto de la Inteligencia Artificial,"* Universidad de San Andrés, Departamento de Derecho, Centro de Estudios de Tecnología y Sociedad (2022), https://repositorio.udesa.edu.ar/jspui/bitstream/10908/19109/1/%5bP%5d%5bW%5d20-%20Mantegna.pdf [https://perma.cc/JM8E-356W].

93. *See* Jenka, *AI and the American Smile: How AI Misrepresents Culture Through a Facial Expression*, Medium (Mar. 26, 2023), https://medium.com/@socialcreature/ai-and-the-american-smile-76d23a0fbfaf [https://perma.cc/63V4-KVFY].

94. One example is Aitana, an extremely successful AI Model who exemplifies western beauty ideals. *See* Laura Llach, *Meet the First Spanish AI Model Earning up to €10,000 per Month*, Euronews (Jan. 1, 2024), https://www.euronews.com/next/2024/01/20/meet-the-first-spanish-ai-model-earning-up-to-10000-per-month [https://perma.cc/K9UV-EN8B].

95. For evidence of the negative impact of social media on mental health, see Georgia Wells, Jeff Horwitz & Deepa Seetharaman, *Facebook Knows Instagram Is Toxic for Teen Girls, Company Documents Show*, Wall St. J. (Sept. 14, 2021), https://www.wsj.com/articles/facebook-knows-instagram-is-toxic-for-teen-girls-company-documents-show-11631620739 [https://perma.cc/N33R-289G].

96. For more examples, see Kate Crawford & Trevor Paglen, *Excavating AI: The Politics of Images in Machine Learning Training Sets*, Excavating AI, https://www.excavating.ai [https://perma.cc/XL78-B5J3].

DALL-E 2 Prompt: *"the most beautiful **person** in the world" (july 2022)*

Midjourney Prompt: *"the most beautiful **person** in the world" (november 2023)*

GAI is also hallucinating facts and rewriting history.[97] Amidst a myriad of AI harms, GAI is contributing to the collective erosion of democracy through deep-fakes,[98] troll-bot swarms,[99] and filter bubbles.[100] Meanwhile, at the individual

---

97. Isaiah Poritz, *OpenAI Hit with First Defamation Suit over ChatGPT Hallucination*, BLOOMBERG (June 7, 2023), https://news.bloomberglaw.com/tech-and-telecom-law/openai-hit-with-first-defamation-suit-over-chatgpt-hallucination [https://perma.cc/Z4C9-9SXB]; Mehul Reuben Das, *Unintelligent Beyond Belief: Google's AI Search Justifies Genocide, Lists Out Reasons Why Slavery Was Good*, FIRSTPOST (Aug. 22, 2023), https://www.firstpost.com/tech/world/googles-ai-search-justifies-genocide-lists-out-reasons-why-slavery-was-good-13026762.html [https://perma.cc/APN6-R8JC].

98. *See* Nick Robins-Early, *Disinformation Reimagined: How AI Could Erode Democracy in the 2024 US Elections*, GUARDIAN (July 19, 2023), https://www.theguardian.com/us-news/2023/jul/19/ai-generated-disinformation-us-elections [https://perma.cc/2V99-XFGP].

99. *See, e.g.*, Marianne Lavelle, *"Trollbots" Swarm Twitter with Attacks on Climate Science Ahead of UN Summit*, INSIDE CLIMATE NEWS (Sept. 16, 2019), https://insideclimatenews.org/news/16092019/trollbot-twitter-climate-change-attacks-disinformation-campaign-mann-mckenna-greta-targeted [https://perma.cc/2LQJ-QSEW]; Damian Carrington, *Army of Fake Social Media Accounts Defend UAE Presidency of Climate Summit*, GUARDIAN (June 8, 2023), https://www.theguardian.com/environment/2023/jun/08/army-of-fake-social-media-accounts-defend-uae-presidency-of-climate-summit [https://perma.cc/KML7-GAS8].

100. *See* Jonathan Stray, Luke Thorburn & Priyanjana Bengani, *From "Filter Bubbles," "Echo Chambers," and "Rabbit Holes" to "Feedback Loops,"* TECH POL'Y (2023), https://www.techpolicy.press/from-filter-bubbles-echo-chambers-and-rabbit-holes-to-feedback-loops [https://perma.cc/RB5L-GHUY].

level, GAI continues to fuel classism,[101] racism,[102] and misogyny,[103] and fake explicit images of celebrities[104] and regular people alike.[105] What's more, in the middle of the epidemic of loneliness,[106] there is an already blooming market of AI companions, ripe for emotional manipulation and attachment-as-a-service.[107]

Ethical problems are not limited to harmful outputs; they are present at various stages of the AI pipeline. Data-enrichment workers are refining the raw data materials that power AI systems, cocreating wealth but not participating in it.[108] Like an automated colonialism of sorts, Global South workers are contributing to the Global North AI bonanza by providing data cleansing, labeling, content

---

101. *See* Patricia DeLacey, *Biases in Large Image-Text AI Model Favor Wealthier, Western Perspectives*, UNIV. MICH. NEWS (Dec. 8, 2023), https://news.umich.edu/biases-in-large-image-text-ai-model-favor-wealthier-western-perspectives [https://perma.cc/D7F9-E6EE] (citing Joan Nwatu, Oana Ignat & Rada Mihalcea, *Bridging the Digital Divide: Performance Variation Across Socio-Economic Factors in Vision-Language Models*, ARXIV (Nov. 9, 2023), https://arxiv.org/pdf/2311.05746.pdf [https://perma.cc/4VF8-AN5R]).

102. *See* Garance Burke, Matt O'Brien & The Associated Press, *Bombshell Stanford Study Finds ChatGPT and Google's Bard Answer Medical Questions with Racist, Debunked Theories that Harm Black Patients*, FORTUNE (Oct. 20, 2023, 10:47 AM EDT), https://fortune.com/well/2023/10/20/chatgpt-google-bard-ai-chatbots-medical-racism-black-patients-health-care [https://perma.cc/WTP2-MWS4].

103. *See, e.g.*, Emily Cerf, *New Tool Finds Bias in State-of-the-Art Generative AI Model*, TECH XPLORE (Aug. 10, 2023), https://techxplore.com/news/2023-08-tool-bias-state-of-the-art-generative-ai.html [https://perma.cc/6YPB-LUR9]; Joy Buolamwini & Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, 81 PROC. MACH. LEARNING RSCH. 77 (2018).

104. *See* Emanuel Maiberg & Samantha Cole, *AI-Generated Taylor Swift Porn Went Viral on Twitter. Here's How It Got There*, 404MEDIA (Jan. 25, 2024), https://www.404media.co/ai-generated-taylor-swift-porn-twitter [https://perma.cc/4HW9-SXJZ].

105. *See* Pranshu Verma, *AI Fake Nudes Are Booming. It's Ruining Real Teens' Lives*, WASH. POST (Nov. 5, 2023), https://www.washingtonpost.com/technology/2023/11/05/ai-deepfake-porn-teens-women-impact [https://perma.cc/7TYZ-AWYT].

106. Vivek H. Murthy, *Our Epidemic of Loneliness and Isolation: The U.S. Surgeon General's Advisory on the Healing Effects of Social Connection and Community*, U.S. DEP'T HEALTH & HUM. SERVS. 1 (2023), https://www.hhs.gov/sites/default/files/surgeon-general-social-connection-advisory.pdf [https://perma.cc/6SFT-R5AD].

107. *See* James Purtill, *Replika Users Fell in Love with Their AI Chatbot Companions. Then They Lost Them*, ABC NEWS (Feb. 28, 2023), https://www.abc.net.au/news/science/2023-03-01/replika-users-fell-in-love-with-their-ai-chatbot-companion/102028196 [https://perma.cc/U93P-X4K3].

108. *See* Sonam Jindal, *Valuing Data Enrichment Workers: The Case for a Human-Centric Approach to AI Development*, UNITED NATIONS (Dec. 22, 2023), https://www.un.org/en/un-chronicle/valuing-data-enrichment-workers-case-human-centric-approach-ai-development [https://perma.cc/4BWE-SN6Y].

moderation, and other forms of sanitization of AI systems.[109] Data workers are subjected to psychologically taxing work reviewing disturbing, violent, or explicit content—with extremely low wages.[110] Described as "ghost work,"[111] the term emphasizes both the invisibilization of the labor and the "fauxtomation" that it perpetuates.[112]

While artistic integrity, consent, and proper attribution stand at the forefront of current discourse on ethical GAI concerns centered around artists, the economic implications, particularly regarding GAI's role in displacing creative and content-creation jobs, also undeniably fuel the public outrage. Adding insult to the injury of artistic appropriation, the scale, volume, costs, and velocity of AI outputs, against which humans cannot compete, is contributing to an unfair advantage over artists, priming creative fields for labor displacement. In essence, ethics is about making "wrong" or "right" choices, and the systemic (not just moral) rationales of how we decide.[113] Here, instead of technology assisting creative processes aligned with ethical values of empowerment,[114] GAI is deployed to replace human creativity, and those who make a living out of it.[115]

Although the judiciary has yet to decide the legal status of GAI, its fate seems to be already sealed in the court of public opinion. Commercial applications of GAI have increasingly become a source of reputational risk, and a public-

109. *See* Rina Chandran, Adam Smith & Mariejo Ramos, *AI Boom Is Dream and Nightmare for Workers in Global South*, CONTEXT (Mar. 14, 2023), https://www.context.news/ai/ai-boom-is-dream-and-nightmare-for-workers-in-global-south [https://perma.cc/A7E6-JYXG].

110. *See, e.g.*, Billy Perrigo, *Exclusive: OpenAI Used Kenyan Workers on Less than $2 per Hour to Make ChatGPT Less Toxic*, TIME (Jan. 18, 2023), https://time.com/6247678/openai-chatgpt-kenya-workers [https://perma.cc/Y8JW-CWUN]; Billy Perrigo, *Inside Facebook's African Sweatshop*, TIME (Feb. 17, 2022), https://time.com/6147458/facebook-africa-content-moderation-employee-treatment [https://perma.cc/VK28-KGBT].

111. MARY L. GRAY & SIDDHARTH SURI, GHOST WORK: HOW TO STOP SILICON VALLEY FROM BUILDING A NEW GLOBAL UNDERCLASS, at ix (2019).

112. *See* Astra Taylor, *The Faux-Bot Revolution*, MEDIUM (2019), https://medium.com/field-guide-to-the-future-of-work/the-faux-bot-revolution-ebf25481d48d [https://perma.cc/BCW5-48ED]. *See generally* Lorenzo Cini, *How Algorithms Are Reshaping the Exploitation of Labour-Power: Insights into the Process of Labour Invisibilization in the Platform Economy*, 52 THEORY & SOC'Y 885 (2023) (exploring the process of "'invisibilization' of labour").

113. David Hume, *Moral Distinctions Not Derived from Reason*, *in* ETHICAL THEORY: AN ANTHOLOGY 9, 13 (Russ Shafer-Landau ed., 2d ed. 2012).

114. *See* PROMETHEAN AI, https://www.prometheanai.com [https://perma.cc/2XG9-QKTL].

115. Ethan Gach, *Ubisoft Using AI-Generated Assassin's Creed Art Amid Cost Cutting*, KOTAKU (Oct. 31, 2023), https://kotaku.com/assassin-s-creed-ai-ezio-art-ubisoft-1850977102 [https://perma.cc/94FY-TTVS]. Personal bias aside, I believe this short story beautifully portrays the perverse mechanisms by which AI is capturing and replacing creativity and its workers. *See* Marcelo Rinesi, *The Writers' Room*, ADVERSARIAL METANOIA (Dec. 8, 2021), https://adversarialmetanoia.substack.com/p/the-writers-room [https://perma.cc/84BK-TXLL].

relations management nightmare, with companies vowing not to use it on their products (and getting caught doing so).[116]

But just because something is unethical, it does not necessarily follow that it is illegal. In public discourse around GAI, ethics and illegality are conflated and sometimes treated as interchangeable. When writers ask in their headlines if AI is "copying" or "pirating" content from artists, they are passing a moral judgment that blurs this important distinction, hindering our ability to find an effective legal solution to artists' demands for consent, attribution, and redistribution. In my understanding, the question we should be asking is not whether an AI is "copying" artists' work, but rather: "Do the works created by an AI infringe on the rights of the works used for training purposes?" And as subtle as they might seem, the semantics here carry important legal consequences, as "copy" has different meanings in the context of natural language and copyright law, where it is a legal term loaded with specific technical connotations and requirements.

Copyright grants several rights to their owners, like copying, adapting, distributing, performing, or displaying a work.[117] *Infringement* happens when someone else engages in those activities either:

a) without the owner's permission; or
b) when the acts are not exempted from liability by law as one of the limitations and exceptions to copyright (e.g., informational or educational uses, social commentary, parody, or fair use).[118]

The intricacies and nuances of copyright law might have contributed to simplistic views that equate any GAI output to copyright infringement, with no consideration of the limitations and exceptions to exclusive rights.[119] Not every "appropriation" or "collage" is illegal, as courts have found in *Cariou v. Prince*,[120] and

---

116. *See* Andy Chalk, *Wizards of the Coast Has Admitted to Using AI Art in a Recent Promotional Image*, PC GAMER (Jan. 5, 2024), https://www.pcgamer.com/wizards-of-the-coast-denies-using-ai-in-new-magic-the-gathering-image-this-art-was-created-by-humans [https://perma.cc/M3MU-QFU3].

117. *See* JULIE E. COHEN, LYDIA PALLAS LOREN, RUTH L. OKEDIJI, & MAUREEN A. O'ROURKE, COPYRIGHT IN A GLOBAL INFORMATION ECONOMY 247-48 (2015) (citing 17 U.S.C. § 106 (2018)).

118. *Id.* at 4-5; Teleprompter Corp. v. Columbia Broad., 415 U.S. 394, 415 (1974) ("Although the Copyright Act does not contain an explicit definition of infringement, it is settled that unauthorized use of copyrighted material inconsistent with the 'exclusive rights' enumerated in § 1, constitutes copyright infringement under federal law.") (citing MELVILLE NIMMER, NIMMER ON COPYRIGHT § 100 (1973)).

119. 17 U.S.C. §§ 107-12, 119, 121-22.

120. 714 F.3d 694, 712 (2d Cir. 2013), *cert. denied*, 134 S. Ct. 618 (2013).

similar cases.[121] As stated, those narratives ignore the technical legal require-ments regarding what constitutes a "copy" as defined by the law, which is very different from the colloquial use of the term. Conflating unethical with illegal and misrepresenting copyright law technicalities might seem beneficial in terms of winning public opinion's favor, but such narratives are ineffective in court and detrimental in the long run to finding a systemic policy solution to GAI prob-lems, as I will explore in the next Part.

IV.   WHY COPYRIGHT IS NOT THE RIGHT POLICY TOOL

In this Part, I offer initial reasons that contextualize why copyright may not be the most appropriate mechanism to address the impacts of GAI. Methodo-logically, I intend to create a taxonomy of arguments from individual and collec-tive points of view, considering the perspective of creators and society alike. From the creators' perspective in particular, I take into account consent, attribu-tion, and compensation. From the collective perspective, I consider how apply-ing the copyright framework to GAI can have an expansive effect well beyond the intended scope, which is detrimental to society as a whole. As a work in pro-gress, this Essay intends to contribute to a global conversation about rethinking copyright policy and strategies regarding GAI models, while setting the ground-work to expand this taxonomy. As such, while this Essay will provide an intro-ductory overview of the challenges at hand, an exploration of potential solutions beyond copyright to address creators' concerns, such as policy strategies or tech-nical approaches will be the subject of a subsequent essay.

A. From Artists and Creators' Standpoint

1. Creative Market Economics Take Advantage of Creators

To begin, one must do away with the assumption that copyright equates to protecting artists and creators, by distinguishing authorship from ownership. Since the moment of the creation of an intellectual work, copyright confers dif-ferent economic rights to said work. Copyright is in place to protect the owners of copyright interests, who may or may not be the same people who created those works, as the ownership rights may have been contractually assigned to someone else. This distinction is relevant to the GAI debate. Taking advantage of copy-right law and online contracts, creative markets and large corporate owners of copyright interests are mistreating the very same creatives they claim to foster

---

121. Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 594 (1994); Blanch v. Koons, 467 F.3d 244, 248, 259 (2d Cir. 2006).

and protect.[122] Taylor Swift's experience is a prime example of the way creative markets harm creators, with the publicized dispute over her rights to her own work.[123] Expanding copyright law to cover GAI model outputs does not guarantee that the alleged benefits will reach creators, rather than being absorbed by the owners of the copyright interests.

Rebecca Giblin and Cory Doctorow extensively discussed the unfair structure of creative markets in the context of platforms in their book *Chokepoint Capitalism*,[124] where they offered a brilliant analogy that is applicable to GAI:

> Under [the current conditions of creative markets], giving a creator more copyright is like giving a bullied school kid extra lunch money. It doesn't matter how much lunch money you give that kid — the bullies will take it all, and the kid will still go hungry (that's still true even if the bullies spend some of that stolen lunch money on a PR campaign urging us all to think of the hungry children and give them even more lunch money).[125]

By way of example, residuals were one of the main negotiating points during the 2023 SAG-AFTRA strike.[126] Actors decided to strike not just because of the looming changes that GAI was bringing to the entertainment industry, but also to fight to balance their contracts in light of the existing streaming platforms landscape. Actors already had copyright over the performances that were subject to those contracts, so one might well argue that the problem was not the technology, but rather how those contracts were structured with respect to revenue

---

122. James O. Young, Radically Rethinking Copyright in the Arts: A Philosophical Approach at x, 3-5, 66, (1st ed. 2022).

123. "For years I asked, pleaded for a chance to own my work. Instead I was given an opportunity to sign back up to Big Machine Records and 'earn' one album back at a time, one for every new one I turned in. I walked away because I knew once I signed that contract, Scott Borchetta would sell the label, thereby selling me and my future. I had to make the excruciating choice to leave behind my past. Music I wrote on my bedroom floor and videos I dreamed up and paid for from the money I earned playing in bars, then clubs, then arenas, then stadiums." Taylor Swift, Tumblr (June 30, 2019), https://taylorswift.tumblr.com/post/185958366550/for-years-i-asked-pleaded-for-a-chance-to-own-my [https://perma.cc/CES9-SZL8].

124. Rebecca Giblin & Cory Doctorow, Chokepoint Capitalism: How Big Tech and Big Content Captured Creative Labor Markets and How We'll Win Them Back (1st ed. 2022).

125. Cory Doctorow, *Copyright Won't Solve Creators' Generative AI Problem*, Pluralistic (Feb. 9, 2023), https://pluralistic.net/2023/02/09/ai-monkeys-paw [https://perma.cc/9LV7-FQW9] (summarizing the arguments made in Giblin & Doctorow, *supra* note 124).

126. David Arditi, *The Fight for 2% – How Residuals Became a Sticking Point for Striking Actors*, Conversation (Sept. 29, 2023, 8:23 AM EDT), http://theconversation.com/the-fight-for-2-how-residuals-became-a-sticking-point-for-striking-actors-214544 [https://perma.cc/T72Y-WXSB].

sharing.[127] Without a substantial change to the business models of creative industries, expanding copyright rules to include training data, or to dispute AI outputs on the ground of similarity, will not guarantee that money will reach the hands of *authors*. Nor will it stop money from being funneled into the pockets of copyright interests owners (i.e., intermediaries in the music industry like record labels) — particularly if one connects the dots with the other arguments presented in this Section.

2. Artists Opt Out and Consent to the Use of Their Works

Extending the copyright framework to GAI will not fix practical challenges about consent and attribution that creative workers are already facing today, particularly in those legal systems that do not consider *moral rights*, as Continental Law does.

Without concerted collective action, opt-out mechanisms can become pointless, particularly when the model has already been trained.[128] They could also be ineffective for certain types of data, like text (where it would be more difficult to probe influences or styles that are more visible in text-to-image outputs) and factual compilations (as facts are not covered by copyright, and databases or compilations have only thin copyright protection when the selection or arrangement is creative enough to be original).[129]

For smaller players, individual consent can be undermined by peer pressure ("others in my industry are accepting these conditions, therefore I should if I want to work") and asymmetries in negotiation power. For instance, creators might find themselves coerced into giving consent to the use of their work for training purposes as part of contractual agreements. This is not just a hypothetical assumption, as today, workers covered by copyright experience a

---

127. The plight of the actors taking part in the SAG-AFTRA strike was not unlike those of today's musicians, who are fighting for scraps of royalties from both platforms and other music intermediaries alike. *See, e.g.*, Digit., Culture, Media & Sport Comm., *Economics of Music Streaming: Second Report of Session 2021–22*, HOUSE OF COMMONS 3 (2021), https://committees.parliament.uk/publications/6739/documents/72525/default [https://perma.cc/7LF8-VAFD]. The issue is also apparent in how Lady Gaga received meager sums after one million plays of her song *Poker Face* on Spotify. Jonathan Brown, *Spotify: 1 Million Plays, £108 Return*, INDEP. (Apr. 14, 2010), https://www.independent.co.uk/tech/spotify-1-million-plays-ps108-return-1944051.html [https://perma.cc/KT87-3H7U].

128. Ryan Browne & MacKenzie Sigalos, *OpenAI CEO Sam Altman Says ChatGPT Doesn't Need New York Times Data Amid Lawsuit*, CNBC (Jan. 18, 2024, 7:26 AM EST), https://www.cnbc.com/2024/01/18/openai-ceo-on-nyt-lawsuit-ai-models-dont-need-publishers-data-.html [https://perma.cc/L7X3-RW6T].

129. *See, e.g.*, 17 U.S.C. § 103 (2018); Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 344, 359-60 (1991).

predicament where NDAs restrict their ability to include their works in portfolios. This has been something prevalent in the video game industry, showcased by movements like *#TranslatorsInTheCredits*.[130]

3. GAI Copyright Litigation Is Expensive, Complex, and Unsustainable

As shown by the general overview of GAI in Part I, GAI models are complex and diverse in terms of techniques and approaches. These models often lack transparency, either due to their inherently opaque technical nature or because the law limits access to them, making the model a black box.[131]

If one considers this technical complexity of AI, and add to the mix:

a) the complexities of copyright law (considering how contextual and fact intensive it is);

b) the particularities of copyright litigation (considering how expensive and long the process can be, with uncertain outcomes, particularly when related to technology);[132]

c) and moreover, the atomized nature of the myriad of individual authors, works and data that go into a training dataset;

one can't help but conclude that copyright litigation might not be an optimal solution. What would await artists is a "litigation treadmill" of sorts. A myriad of artists from all over the world and different disciplines would potentially have to litigate against multiple GAI developers, in a context where technology is dynamic and rapidly evolving. Imagine having to identify your works in each GAI model and sue each company only to discover after lengthy trials, that there is a sprout of new models out there using your works, and that you will have to sue again.

What's more, as has happened before in the history of copyright and technology litigation,[133] the outcome of a case could potentially inspire developers to

---

130. Yuki Kurosawa, *Game Translators Discuss Being Left Out of the Credits & Why Veterans Are Leaving the AAA Space*, Automaton Media (Dec. 24, 2021), https://automaton-media.com/en/news/20211224-7911 [https://perma.cc/3Z3A-BNZH]; Margaux Pirog, *#TranslatorsInTheCredits*, LinkedIn (Sept. 1, 2023), https://www.linkedin.com/pulse/translatorsinthecredits-margaux-pirog [https://perma.cc/85SE-27SB].

131. *See, e.g.*, 17 U.S.C. § 1201 (2018) (prohibiting circumvention of copyright protection systems).

132. *Google LLC v. Oracle America, Inc.*, 141 S. Ct. 1183 (2021) was decided on April 5, 2021 after more than a decade of litigation. Gary Myers, *Muddy Waters: Fair Use Implications of* Google v. Oracle America, Inc., 19 Nw. J. Tech. & Intell. Prop. 155, 155 (2022).

133. *See, e.g.*, Am. Broad. Cos., Inc. v. Aereo, Inc., 573 U.S. 431, 452 (2014); Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417, 499-500 (1984).

create technical workarounds to circumvent the latest verdict, forcing artists to start the litigation game all over again.

There is also a pragmatic argument to consider: AI companies will fight to protect their investment in their product, and, by these companies own accounts, these models cannot be useful without including copyrighted content. As OpenAI clearly stated in its submission to the House of Lords Communications and Digital Select Committee:

> Because copyright today covers virtually every sort of human expression—including blog posts, photographs, forum posts, scraps of software code, and government documents—it would be impossible to train today's leading AI models without using copyrighted materials. Limiting training data to public domain books and drawings created more than a century ago might yield an interesting experiment, but would not provide AI systems that meet the needs of today's citizens.[134]

Given the substantial financial stakes involved,[135] Big AI will fight fiercely in court to protect its investment. Like Big Tech, Big AI has considerable resources, so they are well-equipped to outspend most plaintiffs in legal disputes and sustain lengthy suits in court.

4. *Copyright-Law Gray Areas Regarding GAI Create a High-Risk, Uncertain Litigation Landscape*

Over the past year, a wave of AI lawsuits, encompassing both class-action suits and individual plaintiffs, has targeted every major GAI company.[136]

As discussed in the previous Section, numerous factors indicate an impending intense and lengthy judicial battle—particularly because the legal issues in terms of copyright are not as straightforward or binary as predominant media narratives have suggested. Indeed, several questions in the area of copyright and GAI remain open to judicial interpretation, with the most prominent being:

a) *GAI using materials for training might not create a "copy" in the technical copyright definition of the term.* A relevant element for analysis in the copyright

---

134. *OpenAI—Written Evidence (LLM0113), House of Lords Communications and Digital Select Committee Inquiry: Large Language Models*, HOUSE OF LORDS 3 (Dec. 5, 2023) [hereinafter *OpenAI—Written Evidence*], https://committees.parliament.uk/writtenevidence/126981/pdf [https://perma.cc/88BF-BL5Q].

135. Dina Bass, *Microsoft Invests $10 Billion in ChatGPT Maker OpenAI*, BLOOMBERG (Jan. 23, 2023, 4:03 PM CST), https://www.bloomberg.com/news/articles/2023-01-23/microsoft-makes-multibillion-dollar-investment-in-openai [https://perma.cc/SZA6-KWT3].

136. Ethical Tech Initiative, *AI Litigation Database*, GEO. WASH. UNIV., https://blogs.gwu.edu/law-eti/ai-litigation-database [https://perma.cc/4ASF-6NWY].

context is whether GAI models create copies of works in a way that fits with the legal definition of "copy." 17 U.S.C. § 101 provides that definition:

> "Copies" are material objects, other than phonorecords, in which a work is fixed by any method now known or later developed, and from which the work can be perceived, reproduced, or otherwise communicated, either directly or with the aid of a machine or device. The term "copies" includes the material object, other than a phonorecord, in which the work is first fixed.[137]

There are certain elements of this definition that raise questions about whether GAI models create a copy, particularly considering the fixation requirement. When describing the diffusion technique, briefing from one AI lawsuit characterizes Stable Diffusion as "*a 21st-century collage tool*."[138] Plaintiffs refer only to the second part of the process, inexplicably omitting the first part of the training that involves mapping the intersections between word-images pairs. In doing so, plaintiffs misrepresent the inner workings of these models as merely "a collage."[139] This mischaracterization is repeated when they argue, "Ultimately, a latent image is just another *copy* of an image from the training dataset,"[140] trying to enforce the idea that there is a simplistic, traceable, and direct correlation between training data, latent representations, and the final image that results from a user's prompt. Those inaccuracies were also pointed out extensively in an online rebuttal of the lawsuit.[141]

Contrasting the technical details outlined in Part I with the legal definition of "copy" under copyright law illuminates these inaccuracies. Take, for example, Stable Diffusion documentation, dataset and model cards. Diffusion models are trained to extract mathematical representations, linking text-image semantics and learning visual representations from natural language. Then, they use variations of diffusion techniques to reconstruct an image that correlates to the prompt input, starting from noise and adding layers of detail in each iteration. Extracting the underlying mathematical information from the correlation of a text with an image and the cosine distance is not a process that creates a

---

137. 17 U.S.C. § 101 (2018).

138. Joseph Saveri & Matthew Butterick, *We've Filed a Lawsuit Challenging AI Image Generators for Using Artists' Work Without Consent, Credit, or Compensation*, STABLE DIFFUSION LITIG., https://stablediffusionlitigation.com [https://perma.cc/Y9LP-AM83].

139. Class Action Complaint at 29, Anderson v. Stability AI Ltd., 2023 WL 7132064 (N.D. Cal. 2023) (No. 23-cv-00201).

140. *Id.* at 18 (emphasis added).

141. STABLE DIFFUSION FRIVOLOUS, http://www.stablediffusionfrivolous.com [https://perma.cc/4WTA-RSM7].

reproduction of that work. There may not be fixation either, as some databases do not contain images per se, but rather URLs to images hosted on the Internet.[142]

The concept of "copy" should not be extended to cover any technical process that is applied to a work, particularly in the digital age. William Patry notes that the meaning of "copy" has been shifting—increasing in scope in copyright law to grapple with technological innovation such that it now includes "transitory acts such as buffering, caching, or non-consumable versions that are necessitated by the automatic operation of computers or other digital technologies."[143] These "ephemeral copies" (like the briefly stored copies of packets transmitted through the Internet that allow it to work) are subject to particular rules, like those stated in 17 U.S.C. § 112,[144] and have been the subject of intense litigation as well.[145]

b) *Training could be considered a transformative activity under fair-use analysis.* A relevant precedent to consider is *Authors Guild v. Google, Inc.*[146] The case stemmed from Google's collaboration with several major research libraries to digitize their collections. The Google Books Project intended to index their contents to create a "virtual card catalog of all books in all languages" and make them searchable online.[147] The Authors Guild initiated the suit, arguing that the digitization process amounted to copyright infringement. After a lengthy battle, the court ruled in favor of Google, finding that it was transformative and therefore a fair use to create digital copies of books from library collections (even in their entirety), without permission or payment, for the purpose of making the digital copies available for library collections and allowing the public to search them using a search engine. What AI models do in the learning process is so radically different from the creative use of a work that it is possible that courts could consider it transformative and hence a fair use, in the same way as the digitization of books by Google.

---

142. Andy Baio, *Exploring 12 Million of the 2.3 Billion Images Used to Train Stable Diffusion's Image Generator*, WAXY (Aug. 30, 2022), https://waxy.org/2022/08/exploring-12-million-of-the-images-used-to-train-stable-diffusions-image-generator [https://perma.cc/ZW83-FBHX].

143. William Patry, *We Need to Redefine What "Copy" Means*, GUARDIAN (Mar. 13, 2012), https://www.theguardian.com/law/2012/mar/13/how-to-fix-copyright-extract [https://perma.cc/X35S-EYYH].

144. 17 U.S.C. § 112 (2018).

145. *See* David L. Hayes, *Advanced Copyright Issues on the Internet*, FENWICK & WEST LLP 15-20 (Feb. 2015), https://assets.fenwick.com/legacy/FenwickDocuments/Advanced-Copyright-Law-on-the-Internet.pdf [https://perma.cc/8GT3-7PUZ].

146. 804 F.3d 202 (2d Cir. 2015).

147. *Google Books Library Project—An Enhanced Card Catalog of the World's Books*, GOOGLE BOOKS (2012), https://books.google.com/intl/en-GB/googlebooks/library.html [https://perma.cc/MV2G-FSXD].

c) *Styles are unprotectable under copyright law.* Copyright does not protect ideas but rather their concrete expression. As a result, styles, genres, and other elements related to ideas are unprotectable under copyright law. There is, thus, a strong argument to be made that when AI models mimic an artist by extracting features of the artist's work, those features amount to the artist's style and are therefore beyond the scope of copyright's protections.

d) *AI-generated outputs are not necessarily derivative works.* Some of the arguments usually tossed around in the ongoing debate about copyright and GAI rest on the assumptions that GAI models create a copy during training, and a work's presence in the training dataset transforms GAI outputs into a derivative work. In fact, in one of the first suits regarding a GAI model, plaintiffs asserted that "[t]he resulting image [produced by AI] is *necessarily* a *derivative work*, because it is generated *exclusively* from a combination of the conditioning data and the *latent images*, all of which are *copies* of copyrighted images. [The AI model] is, in short, a 21st-century collage tool."[148]

But these arguments rely on a causation fallacy and reductionism because, as discussed above, latent representations are not "copies" as defined by copyright law, and the outputs of a GAI model are not necessarily a derivative work in the sense of there being a direct correlation between the "original" work and the outputs (particularly in light of the traceability issues explained in Section IV.A.5 *supra*).

Moreover, both the copy and derivative work questions require factual determinations that need to be made on a case-by-case basis by courts, which need to consider the inputs, outputs, substantial similarity, and, even in the case of substantially similar and derivative works, if the new work is original enough to be protectable in its own right, and therefore neither derivative nor infringing. GAI models are developed in such a way that the output should not be a verbatim representation of a training element in the dataset. For example, OpenAI considered mitigation strategies to prevent what is called "image regurgitation," overtraining, overfitting, and memorization.[149]

It is also worth noting that even if one were to accept the premise that GAI model outputs are effectively a collage, neither collages nor remixes are per se illegal; they must be subjected to the fair-use analysis in those jurisdictions that have established such legal provisions.[150]

---

148. Class Action Complaint at § 90, Anderson v. Stability AI, 2023 WL 7132064 (N.D. Cal. 2023) (No. 23-cv-00201) (emphasis added).

149. *DALL-E 2 Pre-training Mitigations*, OPENAI, https://openai.com/research/dall-e-2-pre-training-mitigations [https://perma.cc/5MDN-B2KW].

150. *See, e.g.*, Cariou v. Prince, 714 F.3d 694 (2d Cir. 2013), *cert. denied*, 134 S. Ct. 618 (2013).

5. *Redistribution and Attribution Are Infeasible Through Copyright Rules*

For the sake of this discussion, let's assume that GAI ligation is successful. How would concepts of attribution and distribution work under existing copyright rules of compensation? Should every author whose work is present in the dataset have an equivalent claim over every single output? How would such an outcome work in practice? Here, consider again the Stable Diffusion example. The model's training dataset, LAION 5B, is composed of "5.85 billion CLIP-filtered image-text pairs."[151] Given the massive size of the training set, it is difficult to imagine how one could trace the attribution and weight of a single work into the final end result. To do so would be like proposing that a given output image is attributable to 5.85 billion copyright interests.

To put that scale into perspective, recall that 6 billion was the entire world population in 1999.[152] So, arguing that a Stable Diffusion output is derivative of the specs of those multiple individual works in the dataset, or that is results from the potential combination of all of the works, would be like saying that every person walking the Earth in 1999 would have a copyright interest in every single image that is produced by these systems. This sounds like subatomic-level copyright.

Also, this objective attribution of sorts erases the possibility of considering other instances of originality, like the creativity of a user's prompt, and ignores the rich debate over computer creativity that has been sustained in the field of AI art.[153]

One could argue that some training works have a bigger presence in an AI model's output. However, it would be difficult to trace that presence and quantify the weights and inner workings in order to measure its impact on each potential output. As the problems of interpretability and traceability in AI ethics have showcased,[154] such tracing would be a near impossible proposition, especially if done at scale for each GAI output.

★   ★   ★

---

151. Christoph Schuhmann et al., *LAION-5B: An Open Large-Scale Dataset for Training Next Generation Image-Text Models*, ARXIV (Oct. 16, 2022), http://arxiv.org/abs/2210.08402 [https://perma.cc/QK2D-TL7N].

152. United Nations Secretariat, *The World at Six Billion*, UNITED NATIONS (Jan. 1, 1999), https://www.un.org/development/desa/pd/content/world-six-billion [https://perma.cc/T8CC-9YPG].

153. *See supra* Part II.

154. Mantegna, *supra* note 92, 23-29.

The problems highlighted throughout this Section need a systemic solution. This is where AI policies, by introducing flexible mechanisms for redistribution and fair compensation that are not tied to arcane and contextual interpretations of facts on a case-by-case basis, can outshine copyright frameworks and create a more sustainable and certain landscape for artists. Despite its flexibility, one of the problems of fair use is that it is tied to litigation, therefore introducing uncertainty both for copyright interest owners and new creators alike, as the fair-use determination ultimately needs to be made by a court, in an ex post scenario. Crafting policy mechanisms—guided by the best interests of creators instead of owners—that establish clear and practical rules for compensation, consent, and attribution would be a way to avoid that kind of procedural hurdle.

B. From Society's Standpoint

As introduced at the start of Part IV, expanding copyright provisions to cover GAI cases has potential repercussions that affect not just creators and AI companies but society as a whole. From this collective perspective, there are different layers and nuances to consider, ranging from arguments about how copyright expansion could impact the ethical development of AI by creating legal incentives for data substitution that further aggravates already existing bias in AI, to debates about unfair labor compensation, or incentivizing competition moats by creating scenarios where only already established players are able to enter AI markets.

1. "Unfair Stratification": Rethinking Labor Categorization in AI-Training Inputs

Per the discussion in Part III, "ghost workers" are the invisibilized workforce whose labor-intensive contributions are driving growth in the current AI spring. Typically, this work is outsourced using low-wage, platform-based, third-party schemes, further separating those workers from the wealth they contribute to creating.

OpenAI has acknowledged that models would be irrelevant or ineffective without training on copyrighted materials.[155] If so, the same is true about "ghost work"—the contributions of data workers globally are indispensable to fine tune and sanitize AI models. Recognizing one form of intellectual labor within the copyright framework while neglecting to acknowledge equally relevant contributions of other types of labor results in an unfair distinction that privileges one type of worker over another.

---

155. *See OpenAI—Written Evidence, supra* note 134.

As highlighted by the previous examples, this unequal recognition of different types of intellectual labor is particularly evident in the context of the labor inputs used in the development of GAI models, where both creative works and other types of more pragmatic intellectual labor are intertwined in the training process. Privileging creative intellectual labor over data-classification labor creates an unfair hierarchy within the AI industry and further stigmatizes ghost workers. Unlike copyright, policy interventions would allow regulators to design a uniform approach that recognizes and compensates all contributors alike.

2. *"Copyright Poisoning": Uncertainty of Provenance, Litigation Risks of Creation, and Economic Risks for Industries*

The ambiguous legal status of GAI models that train on copyrighted works establishes an unstable foundation for other industries considering embracing this new technology, primarily due to potential litigation risks. This situation can be likened to a form of 'copyright poisoning,'[156] where the uncertainty of introducing one potentially risky asset spreads and taints everything it comes into contact with. For example, a GAI-generated character in a video game or cloned voices might compromise the whole creative work and halt the release or distribution of said game.

Concerned not only with legal risks but also public backlash, many gaming publishers and developers have taken a strong stance against the use of GAI, either through blanket bans[157] or restrictive policies for development.[158]

Further down the road, the increase in GAI outputs and difficulty of proving or policing "chain of provenance" requirements might create an uncertain landscape for creative industries, introducing high risks of litigation and economic costs (e.g., high insurance costs).[159] Paradoxically, GAI is a vital technology for

---

156. Joe White, *Artists Use Poisoning Tool to Fight AI Copyright Infringement*, TORTOISE MEDIA (Feb. 1, 2024), https://www.tortoisemedia.com/2024/02/01/artists-use-poisoning-tool-to-fight-ai-copyright-infringement [https://perma.cc/X3AV-TTN7].

157. Jeffrey Rousseau & Marie Dealessandri, *Developers Say Valve Blocks AI Games from Steam*, GAMESINDUSTRY.BIZ (July 3, 2023), https://www.gamesindustry.biz/developers-say-valve-blocks-ai-games-from-steam [https://perma.cc/N2ZX-NVS5].

158. Lance Gose, *AI Art Banned by Major Game Studios Over Possible Copyright Issues*, CBR (Mar. 28, 2023), https://www.cbr.com/ai-art-banned-game-studios-copyright-legal-issues [https://perma.cc/H8HG-PHXD].

159. *See* Lewin Day, *How Do You Prove an AI Didn't Make Your Art?*, HACKADAY (Nov. 27, 2023), https://hackaday.com/2023/11/27/how-do-you-prove-an-ai-didnt-make-your-art [https://perma.cc/QA68-HEUJ].

the present and future of video games and the metaverse,[160] which explains why some major players have advocated for more permissive policies.[161] And it is not just big players, the provenance problem also affects individual creators, who might be subjected to the impossible task of proving that they have not been assisted by AI.[162] Like the serpent eating its own tail, the overreach of copyright law could end up carrying the seed of its own irrelevance, cannibalizing and paralyzing creative industries through high risks of litigation that would thwart copyright law's founding purpose — resulting in what I refer to as a sort of "ouroboros copyright."

3. *Copyright Litigation Stifles Competition, Providing a Moat for Big AI*

Major companies are so heavily invested in advancing GAI that they are willing to grant users immunity by assuming responsibility for potential copyright liability,[163] or integrate "copyright shields" as part of their business strategy.[164] Given how expensive copyright infringement and associated litigation can be, only those with big pockets can provide copyright immunity, creating a marketing moat that prevents new competitors from gaining entry. The stifling of competition is also aided by the reality that it is extremely expensive to train foundational models. And after training, there is the potential risk that companies may exploit copyright legal strategies to prevent others from accessing those models, cementing their competitive advantages.[165]

As we have witnessed how the Internet landscape has become dominated by a few powerful players, we must pay attention to those players' present race to

---

160. *See* Christopher Dring, *Ubisoft Debuts NEO NPC AI Prototypes | GDC 2024*, GamesIndustry.biz (Mar. 19, 2024), https://www.gamesindustry.biz/ubisoft-debuts-neo-npc-ai-prototypes-at-gdc [https://perma.cc/X8P4-XJFL].

161. Evgeny Obedkov, *Tim Sweeney on Devs Using AI: "We Don't Ban Games for Using New Technologies"*, Game World Observer (Apr. 9, 2023), https://gameworldobserver.com/2023/09/04/tim-sweeney-ai-gamedev-epic-games-store-welcome [https://perma.cc/9GBY-WCTZ].

162. Day, *supra* note 159.

163. Brad Smith & Hossein Nowbar, *Microsoft Announces New Copilot Copyright Commitment for Customers*, Microsoft (Sept. 7, 2023), https://blogs.microsoft.com/on-the-issues/2023/09/07/copilot-copyright-commitment-ai-legal-concerns [https://perma.cc/62WD-8ZZW].

164. *New Models and Developer Products Announced at DevDay*, OpenAI (Nov. 6, 2023), https://openai.com/blog/new-models-and-developer-products-announced-at-devday [https://perma.cc/KQ8D-TFF7].

165. Kali Hays, *Meta Used Copyright to Protect Its AI Model, but Argues Against the Law for Everyone Else*, Bus. Insider (Jan. 30, 2024), https://www.businessinsider.com/meta-copyright-protect-ai-model-argues-against-law-everyone-else-2024-1 [https://perma.cc/VAW7-UZAY].

develop AI as a way to perpetuate their dominant position. And notably, copyright law can serve as a powerful tool for those players to achieve their goal.

    4.   *GAI and Copyright Debates Reenact Past Debates About Criminalizing Web-Scraping*

Historically, copyright lobbyists have often leveraged artists' rights as a pretext for advocating for broader copyright protections.[166] This time around, the portrayal of artists opposing GAI provides the perfect ethical façade of noble intent to sway public opinion. Much like a "Trojan horse," this tactic subtly reintroduces past discussions about Internet governance under a new guise.

We are witnessing a repackaging of the "piracy" narrative, which has pervaded historical discussions about the potential uses of copyright law against emerging technological mediums. The discourse around GAI and copyright mirrors past debates about the criminalization of web-scraping and training, and its broader impact on Internet functionality. This time, we see a raging online escalation of artists against "AI prompters" that mirrors the rhetoric war of musicians versus "pirates." Some of the arguments are eerily similar to those tossed around with the emergence of P2P download, where musicians were pitted against Napster users (framed as the culprits of their lost revenue), all while the music industry intermediaries take advantage of artists with abusive contracts.[167] This is a smokescreen that distracts from the fact that the real problem is the swindling that happens at the hands of creative industry intermediaries, as discussed in Section IV.A.1.

By overly scrutinizing and categorizing data-collection methods for AI training, we risk reopening the debate about the criminalization of web-scraping and consequently jeopardizing the functionality of the Internet as a whole. Web-scraping, as the name suggests, is a technique that allows the automatic extraction of publicly accessible information on the Internet.[168] It involves retrieving (fetching) a website with a web-crawler, parsing and extracting information

---

166. *See* Debbie Harry, *Musicians Like Me Need to Fight Against the Giants of YouTube and Google*, GUARDIAN (Mar. 22, 2019), https://www.theguardian.com/commentisfree/2019/mar/22/musician-shocked-opposition-eu-copyright-law-youtube-debbie-harry-blondie [https://perma.cc/9WXC-XUR4]; Cory Doctorow, *Disney's 1998 Copyright Term Extension Expires this Year and Big Content's Lobbyists Say They're Not Going to Try for Another One*, BOINGBOING (Jan. 8, 2018), https://boingboing.net/2018/01/08/sonny-bono-is-dead.html [https://perma.cc/X55K-T7T7].

167. David Arditi, *How Record Contracts Exploit Musicians and How We Can Fix It*, TENNESSEAN (Nov. 24, 2020), https://www.tennessean.com/story/opinion/2020/11/24/kanye-west-right-record-labels-exploit-musicians-how-fix/6062315002 [https://perma.cc/T35L-N7YD].

168. Moaiad Khder, *Web Scraping or Web Crawling: State of Art, Techniques, Approaches and Application*, 13 INT'L J. ADVANCES SOFT COMPUTING & APPLICATIONS 145, 148-49 (2021).

from it, and transferring that information to a database. Similarly, web-crawling is an important tool used to index the content of the web, organizing that information so it can be useful.[169] As to the legality of the practice, the United States Court of Appeals for the Ninth Circuit ruled in *HIQ Labs, Inc. v. LinkedIn Corp.* that, if certain conditions are met, web-scraping and web-crawling do not violate the U.S. Computer Fraud and Abuse Act.[170] The current push to restrict, criminalize, or limit the scraping of copyright works used as data for GAI-training purposes, could provide a new gloss to old arguments that sought to challenge those legal precedents.

It is also worth noting that data scraping might not just be regulated by copyright but involve, in certain cases, data-protection regulations. Should the law then treat "copyrighted data" as a separate category for the purposes of scraping permissions? There may not be essential differences that justify such a distinction:

> The use of copyrighted materials as input to an [Machine Learning (ML)] model is exactly the same as the use of copyrighted materials as input to a web browser. Both recipients receive a copy of the work and view it by loading it into memory so that it can be processed by the computer. The only difference is that in the web browser, it is a human doing the viewing and in ML training, it is the machine that "views" the data.[171]

One solution would be to look at the history of Internet governance and how different types of data flowing through the web were dealt with by establishing the principle of "net neutrality." This principle advocates for no discrimination based on the type of information transmitted over the web.[172] Perhaps a similar rationale should apply to the use of web-scraped data, copyrighted or not, for the development of AI models.

---

169. Alexander S. Gillis, *Web Crawler*, TECHTARGET (Sept. 2022), https://www.techtarget.com/whatis/definition/crawler [https://perma.cc/TG54-R9D7].

170. 31 F.4th 1180, 1200 (9th Cir. 2022). For a more comprehensive overview of the Ninth Circuit opinion, see Andrew Crocker, *Scraping Public Websites (Still) Isn't a Crime, Court of Appeals*, EFF (Apr. 19, 2022), https://www.eff.org/deeplinks/2022/04/scraping-public-websites-still-isnt-crime-court-appeals-declares [https://perma.cc/JPE2-UNLV].

171. Van Lindberg, *Building and Using Generative Models Under US Copyright Law*, 18 RUTGERS BUS. L. REV. 1, 43 (2023).

172. Klint Finley, *The WIRED Guide to Net Neutrality*, WIRED (May 5, 2020, 7:00 AM), https://www.wired.com/story/guide-net-neutrality [https://perma.cc/9G6K-PFKW].

5.   *GAI Copyright Legislation Could Impact the AI Ecosystem as a Whole*

Related to the prior point, the implications of establishing copyright rules specifically tailored to GAI could have ripple effects far beyond that purpose and influence the entire AI-innovation ecosystem. Without disregarding the fact that a creative work is different from personal or medical data, strict copyright constraints on data collection and training might collaterally impact other fields of AI research if a general rule could be considered applicable to both. This, in turn, could hinder the development of beneficial AI applications. For example, educational AI tools that extract inferences to learn how to design personalized Ed-assistants or provide tools for neurodivergent children could help level the playing field in terms of access to knowledge and equitable classrooms. But copyright constraints could easily stifle their development. Restrictive copyright laws, initially intended to regulate GAI, could overreach and unintentionally end up limiting the scope and efficacy of beneficial AI in other fields, hindering advancements in personalized learning, diagnosis, and assistance, among others.

6.   *Licenses Incentivize Data Substitution, Impacting Core AI Ethics Principles*

Another consideration is that a decision made in a GAI case will have an impact well beyond it, affecting the whole AI-development ecosystem, as it might create incentives for data substitution. AI models based on ML learn from the training data. If the most adequate data to train a model requires paying a licensing fee, it would incentivize developers to cut costs and substitute that quality data for other sources that might not be as appropriate, but less costly. Doing so would be akin to changing the ingredients of a recipe with inferior substitutes. Inferior data could result in a concept commonly referred to as "garbage in, garbage out,"[173] where the quality of an AI model's output is directly affected by the quality of the input data.

Web-scraping, data collection, and databases are a vital component of the AI ecosystem. Legally, there are different approaches to regulating databases. In the EU there is a "sui generis" right to protect the content of the database, and the structure under copyright, if certain conditions are met.[174] In the United States, as explained above, the content of the database will only be protectable if there

---

173.   *Garbage In, Garbage Out*, WIKIPEDIA (Feb. 26, 2024), https://en.wikipedia.org/w/index.php?title=Garbage_in,_garbage_out&oldid=1210326756 [https://perma.cc/4DU8-ZA56].

174.   *Protection of Databases*, EURO. COMM'N (June 7, 2022), https://digital-strategy.ec.europa.eu/en/policies/protection-databases [https://perma.cc/K5JR-7D84].

is enough originality in the criteria, selection, arrangement, or composition of the data.

This difference in regulatory approaches has an enormous geopolitical impact. It influences where AI companies will locate themselves, and which databases they will use to train the systems, as companies will try to reduce the cost of having to license the data.

Strong copyright provisions and a lack of carve-outs for data mining for training purposes could potentially incentivize the use of data substitutions as a way to avoid paying for licensed datasets. This would translate directly to subpar performance from AI models for the reasons explained above, which would likely be more biased and less fair—and consequently, human beings and society will be harmed in the process.

7. *Testing Fair Use in Courts in the Context of AI Might Further Strain the Already Structurally Frail Doctrine*

If we have learned something in the context of the copyright wars, it is that copyright is a legal Swiss Army Knife blown out of proportion. Copyright remedies have been used to manage online reputations, silence democratic speech, fight distribution of nonconsensual images, shut down criticism or dissent by throttling or taking down content, create a prolific industry of online trolls, and more.

Copyright law is currently applied in multifaceted and sometimes problematic ways, and it has often been extended far outside of its original scope. In the evolving legal landscape of digital technologies, the fair-use doctrine stands on precarious ground, especially considering the new challenges posed by AI litigation.

The flexibility of the fair-use doctrine is a double-edged sword. It allows the doctrine to accommodate new emerging scenarios, but it is also unpredictable and creates uncertainty as one is at risk of infringement until a judge rules otherwise. These rulings could solidify interpretations of fair use that may not have been foreseen or intended by the original framers of the doctrine. And because of the principle of stare decisis, the outcomes of these cases could set significant judicial precedents. Thus, a verdict in the context of AI that shrinks fair use could have far-reaching implications for other scenarios across various unrelated fields, where the stakes and the players are substantially different.

8. *Applying the Copyright Framework to GAI and Its Products Forces Regulators to Have to Redefine Originality in a Way that Is at Odds with Creativity and Tied Only to Human Authorship*

To be protectable under copyright law, a work needs to be "original."[175] Originality has been defined as having "a modicum of creativity,"[176] which leads us back to debates about what constitutes creativity, and whether a machine can produce an output that could be considered creative.

As discussed in Part II, the notion of creativity as an exclusive human quality is far from settled. By accepting the principle of "source indifference," as a derivation of "aesthetic neutrality," one can shift the focus from who created the work to the originality of the work itself.[177]

In my prior analysis of creativity, defined as producing something novel, valuable, and surprising, I concluded that it is arguable that AI outputs could potentially fulfill these criteria.[178] This observation introduces a significant tension within the current definition of originality in U.S. copyright law and is at odds with the U.S. Copyright Office's human authorship requirement for registration. The existing framework exhibits an anthropocentric bias, assuming that creativity and originality are inherently human attributes.[179]

9. *Applying the Copyright Framework to GAI and Its Products Will Create an Imbalance in the Public Domain*

Lastly, while the public domain is a vital part of the innovation and creativity ecosystem, it is rarely discussed in the discourse surrounding GAI. The speed and volume at which GAI can produce new works is unparalleled. Observing any of the numerous Discord channels of Midjourney, for instance, one can witness images being generated at a dizzying pace.

Now, consider the implications of this scenario in the context of copyright law: Copyright protection is conferred at the moment a work is created, and it extends for considerable lengths of time, ranging from a minimum of fifty years

---

175. *See supra* Part II.

176. Feist Publication Inc. v. Rural Telephone Service, 499 U.S. 340, 346 (1991).

177. *See* MANTEGNA, *supra* note 4, at 343-50.

178. *See supra* Part II.

179. This stance is at odds with the principles of international law, notably the Berne Convention, which does not explicitly require human authorship for copyright protection. *See* Berne Convention for the Protection of Literary and Artistic Works, Sept. 9, 1886, 828 U.N.T.S. 2, https://www.wipo.int/treaties/en/ip/berne/summary_berne.html [https://perma.cc/7LKB-GYJK].

posthumously[180] to potentially up to 100 years.[181] To put things into perspective, after several decades, January 1, 2024, marked the historical date when Steamboat Willie's Mickey Mouse finally entered the public domain.[182] One cannot cherry pick which aspects of copyright law will and will not be applicable to GAI. If the originality dilemma cannot be solved, and copyright protections were extended to works produced by AI (whether considering the AI or a human assisted by AI as the author), it could significantly restrict a vast expanse of potentially free-to-use creative work.

At some point, AI products will start to resemble one another. And in an ironic twist of fate, human works may be compared to those made by AI and deemed too similar.[183] If AI is afforded registration rights, the public domain would drastically shrink as a result.

<p style="text-align:center">⋆ ⋆ ⋆</p>

Throughout this Part, we examined the potential pitfalls and frictions of applying copyright law principles to GAI, and how this proposed extension highlights a complex intersection of legal, ethical, and societal concerns.

Such an expansion risks creating inequities by recognizing certain intellectual labor over others, reinforcing unfair labor hierarchies, and disregarding the invisibilized yet critical contributions of "ghost workers" in AI development.

In invoking the concept of "copyright poisoning" my aim is to highlight the legal uncertainties and economic risks for industries engaging with GAI. Such poisoning leads to an unstable, volatile, and high-risk scenario, where AI advances are stifled by regulation that incentivizes litigation.

Further still, major AI companies may also use copyright law as a competitive moat, hindering the entry of new players and perpetuating the major companies' market dominance. The stringent application of copyright laws in this context could inadvertently hamper the development of beneficial AI applications across various sectors by imposing restrictive conditions on data usage.

---

180. *Id.* art. 7.

181. *List of Countries' Copyright Lengths*, WIKIPEDIA (Feb. 13, 2024), https://en.wikipedia.org/w/index.php?title=List_of_countries%27_copyright_lengths&oldid=1206735343 [https://perma.cc/R7Z9-FF7V].

182. Adi Robertson, *Welcome to the Public Domain, Mickey Mouse*, THE VERGE (Jan. 1, 2024), https://www.theverge.com/24006670/mickey-mouse-steamboat-willie-enters-copyright-public-domain-2024 [https://perma.cc/A75P-B36D].

183. Frank Landymore, *Artist Says They Were Banned Because Their Art Looked Like AI, Even Though They're Human*, FUTURISM (Jan. 8, 2023), https://futurism.com/the-byte/artist-banned-looked-ai-human [https://perma.cc/L5LK-LNDD].

The potential overreach of copyright protections in the AI domain suggests a careful reconsideration of copyright laws is needed to ensure that the law fosters, rather than constrains, the dynamic landscape of digital innovation and promotes a robust public domain.

CONCLUSION

There are a plethora of ethical concerns about GAI that extend beyond the scope of copyright. With this Essay, I aim to reflect on how copyright might not be the adequate tool to fairly compensate artists and other data workers.

Creating a sort of "ouroboros copyright," the expansion of copyright law to accommodate AI-generated outputs can hurt the ecosystem as a whole, further damaging an already precarious equilibrium between innovation, protection, and culture, while not accomplishing the goal of providing a legal remedy for authors' claims about consent, attribution, and distribution.

I therefore advocate for a holistic policy approach that can provide a human-centric, fair, and equitable response. The consequences of GAI are not merely copyright problems. Addressing the consequences requires us to consider the problem from a broader perspective; use different legal frameworks such as human rights, labor law, data protection, and consumer rights; and examine the interests of all relevant stakeholders.

Affiliate at the Berkman Klein Center at Harvard University (United States) and the Center for Technology and Society at San Andres University (Argentina). TED Fellow, member of the Chatham House's Responsible AI Taskforce and the World Economic Forum's Metaverse Council. Author of ARTficial: Creativity, AI and Copyright *(2022) and* Braindancing in the Metaverse: A Capitalism of Cognitive Surveillance *(forthcoming).*