School of Law
UNIVERSITY OF GEORGIA
Prepare. Connect. Lead.

Journal of Intellectual Property Law

Volume 32 | Issue 1 Article 7

March 2025

The Art That Makes the AI Artist: AI's Potential as a Copyright Infringer and Its Future Under a Licensing Requirement

Bailee Jetton
University of Georgia School of Law

The Art That Makes the AI Artist: AI's Potential as a Copyright Infringer and Its Future Under a Licensing Requirement

Cover Page Footnote

J.D. Candidate, 2025, University of Georgia School of Law. Bailee Jetton graduated with Highest Honors from the Georgia Institute of Technology with a B.S. in Literature, Media, & Communications. Bailee would like to thank God for His many blessings, Professor Joseph Miller for his guidance and expertise, and her friends and family for their support.

NOTE

THE ART THAT MAKES THE AI ARTIST: AI'S POTENTIAL AS A COPYRIGHT INFRINGER AND ITS FUTURE UNDER A LICENSING REQUIREMENT

*Bailee Jetton**

*J.D. Candidate, 2025, University of Georgia School of Law. Bailee Jetton graduated with Highest Honors from the Georgia Institute of Technology with a B.S. in Literature, Media, & Communications. Bailee would like to thank God for His many blessings, Professor Joseph Miller for his guidance and expertise, and her friends and family for their support.

I. INTRODUCTION

"When we program a computer to make choices intelligently . . . we must program it to take an attitude towards its freedom of choice essentially isomorphic to that which a human must take to his own."[1] Computer science professor John McCarthy had this to say twenty-three years after he coined the term "artificial intelligence" (AI).[2] As AI companies have continued to expand their systems' capabilities in the twentieth century, Professor McCarthy's vision for machine programming has become paramount to the future of AI. What he perhaps did not know in 1979 was that AI technology would interplay with the internet and raise serious concerns over the protection of intellectual property rights.

From serving as an integrated tool in the workplace to powering a schoolchild's virtual flashcard set, AI programs are quickly becoming commonplace in everyday life.[3] One of the largest generative AI companies boasts over 180.5 million users as of March 2024 and attracts 100 million users every week.[4] For AI companies, a license requirement risks limiting what AIs can be trained with, thus restricting their output capabilities. For copyright owners, the use of their material to train AIs could be considered an unprecedented form of copying that threatens to overtake human creativity in the market. AI's capabilities can rival human authorship and compete with human-generated works.[5] However, if generative AI is permitted to copy the

[1] John McCarthy, *Ascribing Mental Qualities to Machines*, STAN. UNIV. 25 (1979), http://jmc.stanford.edu/articles/ascribing/ascribing.pdf.

[2] Chris Smith, *The History of Artificial Intelligence*, UNIV. OF WASH. 4 (2006), https://courses.cs.washington.edu/courses/csep590/06au/projects/history-ai.pdf.

[3] *See Artificial Intelligence*, SALESFORCE, https://www.salesforce.com/products/einstein-ai-solutions/ (last visited Sept. 3, 2024); *see also Generative AI,* ADOBE, https://www.adobe.com/sensei/generative-ai.html (last visited Sept. 2, 2024); Quizlet, *Quizlet Launches Advanced AI-Powered Tools for Next-Gen Studying*, PR NEWSWIRE (Aug. 8, 2023), https://www.prnewswire.com/news-releases/quizlet-launches-advanced-ai-powered-tools-for-next-gen-studying-301895290.html.

[4] Oskar Mortensen, *How Many Users Does ChatGPT Have? Statistics & Facts (2024)*, SEO.AI (Apr. 24, 2024), https://seo.ai/blog/how-many-users-does-chatgpt-have.

[5] Omar Siddique, *Can Academics Tell the Difference Between AI-generated and Human-Authored Content?*, TIMES HIGHER EDUC. (Feb. 15, 2024), https://www.timeshighereducation.com/campus/can-academics-tell-difference-between-aigenerated-and-humanauthored-content ("We discovered that participants were generally quite bad at discerning between AI-generated and human-authored content. With a score just above chance, participants had a 52 per cent success rate at identifying the given abstract (as AI or human generated) correctly."); Anna Gallegos-Cannon, *Can You Tell the Difference Between Genuine and AI-Generated Art?*, 90.3 WPLN NEWS (Jan. 5, 2023), https://wpln.org/post/can-you-tell-the-difference-between-genuine-and-ai-generated-art/ ("AI-created images are so well done that it can be difficult (for the

work of human creators, then it may ultimately push human creators out of the market and starve itself of the training materials it needs.

II. BACKGROUND

A. THE RESURGENCE IN AI POPULARITY

AI systems have existed for decades, and this is not the first time AI has experienced a spike in popularity.[6] The digitalization of information and the availability of enormous swaths of information have contributed to AI's latest revival.[7] The internet presents one of the largest collections of data for AI programmers, with over 200,000,000 active websites and 175 new ones created every minute.[8] With such a vast amount of information and a constant influx of new sites, the internet can be an endless source of material for training AI systems.[9] Additionally, with companies making open-source AI programs available to the public, AI has become more accessible to the general population.[10] Some critics question whether this current surge in AI popularity will last or if AI is just a "shiny new toy" captivating consumers until its newness wears off.[11] On the contrary, one current survey found that 79% of its respondents spanning seven industries have used AI at least once, and 22% reported using it regularly.[12] These statistics, coupled with AI's rising availability to the general population, make AI's position as the next frontier technology much more likely. This expansion, along with AI's ability to store and use vast amounts of digitized information, has caused concern for human copyright

average person) to tell the difference between a drawing created by the hand of an artist or a computer within seconds.")

[6] *See History of AI: Timeline and The Future*, MARYVILLE U. (May 19, 2023),**Error! Hyperlink reference not valid.** https://online.maryville.edu/blog/history-of-ai/.

[7] SuryaCreatX, *AI: The Catalyst of Change in the 4th Industrial Revolution – Here's Why*, MEDIUM (Aug. 16, 2023), https://suryacreatx.medium.com/ai-the-catalyst-of-change-in-the-4th-industrial-revolution-heres-why-8de4ab50a49d.

[8] NJ, *How Many Websites Are There in the World?*, SITEEFY, https://siteefy.com/how-many-websites-are-there/ (last updated Aug. 28, 2024).

[9] SuryaCreatX, *supra* note 7.

[10] Edd Gent, *The Tech Industry Can't Agree on What Open-Source AI Means. That's a Problem.*, MIT TECH. REV. (Mar. 25, 2024), https://www.technologyreview.com/2024/03/25/1090111/tech-industry-open-source-ai-definition-problem/ ("On the face of it, open-source AI promises a future where anyone can take part in the technology's development. That could accelerate innovation, boost transparency, and give users greater control over systems that could soon reshape many aspects of our lives.").

[11] Erik Urdang, *AI Isn't a Shiny New Toy. Or Is It?*, GORILLA LOGIC (May 19, 2023), https://gorillalogic.com/blog/ai-in-product-development.

[12] Michael Chui et al., *The State of AI in 2023: Generative AI's Breakout Year*, MCKINSEY (Aug. 1, 2023), https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-ais-breakout-year#.

owners. This is particularly an issue with generative AI programs because of their extensive training processes.

B. GENERATIVE AI AND HOW IT WORKS

There are several types of AI programs and even more forms that AI can take within existing systems.[13] Generative AI refers to AI models that generate entirely new content from user input.[14] Generative AI companies are largely using one of two machine learning frameworks: the older generative adversarial network (GAN) and the newer diffusion model.[15] Within a GAN system are two sub-networks called the generator and the discriminator.[16] When a user enters an input for a specific image, the generator creates an output and presents it to the discriminator.[17] The discriminator, true to its name, then classifies the generator's output as either a fake image or a match for what it knows to be a real image.[18] The discriminator knows what is a true match based on an internal compilation of real images uploaded to it by the programmers.[19] The generator's sole goal is to figure out how to convince the discriminator that the generator's output is real.[20] If it fails, it readjusts the output until the image it creates is indistinguishable from the discriminator's collection of "real" images.[21] It involves a complex algorithm that works by analyzing the relevant data in its system, determining patterns within the data, and returning an original output in accordance with those patterns.[22]

The diffusion model, however, relies on the same kind of collection of base material as a GAN but eliminates the generator-discriminator communication.[23] Instead, the entire training process consists of the model creating an image based on characteristic notes programmers attach to each work in its collection.[24] Before the model begins training, all the images in the collection are labeled with

---

[13] *What is Generative AI?*, NVIDIA, https://www.nvidia.com/en-us/glossary/data-science/generative-ai/#:~:text=How%20Does%20Generative%20AI%20Work,generate%20new%20and%20original%20content (last visited Sept. 5, 2024).

[14] *Id.*

[15] Bella Isaacs-Thomas, *How AI Turns Text into Images*, PBS, https://www.pbs.org/newshour/science/how-ai-makes-images-based-on-a-few-words (last updated Jan. 12, 2023, 3:46 PM).

[16] *Id.*

[17] *The Generator*, GOOGLE, https://developers.google.com/machine-learning/gan/generator (last visited Sept. 15, 2024).

[18] *Id.*

[19] *Id.*

[20] *Id.*

[21] *Id.*

[22] *What is Generative AI?*, *supra* note 13.

[23] *Id.*

[24] Isaacs-Thomas, *supra* note 15.

descriptive captions.[25] The model then learns what words are associated with which materials so that, when a user enters a request, the model can return an accurate output.[26]

Generative AI can produce outputs in various types of mediums, including writing and visual art.[27] Notably, generative AI demonstrates output capabilities that can closely replicate human-made works.[28] Generative AI models are able to produce these outputs after undergoing a process that supplies AI with the training data.[29] For GANs and diffusion models, this training data is the compilation of material used for training the neural networks in both models.[30] Some AI applications use internal data sets held by the company, obtained by consumer actions and with their consent.[31] Others rely upon external material gathered from large collection sites, such as the internet.[32] One version of ChatGPT, a language-focused generative AI model, trained on approximately 300 billion words lifted from internet sources.[33] Moreover, image-focused generative AIs are skimming hundreds of thousands of images as one model report noted:

> 67k images from DeviantArt, . . . and 28k from Tumblr. . . . 244k images came from Shopify, 189k each from Wix and Squarespace . . . and just over 47k from Etsy. . . . Unsurprisingly, a large number came from stock image sites. . . . 15k from Getty Images, 10k from VectorStock, and 10k from Shutterstock, among many others.[34]

[25] *Id.*

[26] *Id.*

[27] *What is Generative AI?*, *supra* note 13.

[28] *See Is Artificial Intelligence Set to Become Art's Next Medium?,* CHRISTIE'S (Dec. 11, 2018), https://www.christies.com/en/stories/is-artificial-intelligence-set-to-become-art-s-nex-0cd01f4e232f4279a525a446d60d4cd1 (describing how art made by an AI rivaled human painters at an art auction, making over $430,000).

[29] *What is Generative AI?*, *supra* note 13.

[30] *Id.*

[31] *See Content Analysis FAQ*, ADOBE, https://helpx.adobe.com/manage-account/using/machine-learning-faq.html#CanIturnoffoptoutofmachinelearning (assuring users their content will not be used to exclusively train the AI tool without their consent) (last updated July 30, 2024).

[32] Kevan Schaul et al., *Inside the Secret List of Websites that Make AI Like ChatGPT Sound Smart*, WASH. POST. (Apr. 19, 2023, 6:00 AM), https://www.washingtonpost.com/technology/interactive/2023/ai-chatbot-learning/.

[33] Alex Hughes, *ChatGPT: Everything You Need to Know About OpenAI's GPT-4 Tool*, BBC Science Focus (Sept. 1, 2023, 9:35 AM), https://www.sciencefocus.com/future-technology/gpt-3.

[34] 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/.

Within each of these source groups, there may be tens of thousands of copyrighted materials that require the owner's permission to replicate.[35] As a result, externally trained generative AI systems have become a hub for copyright violation suits in the creative industry.[36]

C. THE COPYRIGHT STATUTE AND FAIR USE DOCTRINE

Several groups have brought lawsuits against AI companies for using copyrighted materials to train AI systems.[37] Under the Copyright Act of 1976, copyright owners have the exclusive right to make copies or reproductions of their work.[38] Anyone who violates these rights may be subject to infringement actions for each instance of unlawful use.[39]

To establish a violation, the copyright holder must prove that the infringer "actually copied" the owner's work and that the infringer's work is "substantially similar" to the owner's work.[40] Courts understand that owners may not have direct evidence of copying against infringers. If this is the case, courts will consider whether the accused "had access to the work" or whether the similarities between the works are so obvious that copying likely took place.[41] The test for determining whether similarities are condemning, however, "is difficult to define and varies across U.S. courts".[42] Courts may, among other tests, take a holistic

[35] *See, e.g.*, *Collection Image Fees*, Getty, https://www.getty.edu/legal/image_request/fees_schedule.pdf (stating that "[u]nless otherwise noted, the release of a digital image by Getty in no way grants or denies copyright permission for reproduction in any publication, production, or media. Identifying copyright holders and obtaining permission is the responsibility of the requester.") (last visited Sept. 15, 2024).

[36] Ina Fried, *Generative AI is a Legal Minefield*, AXIOS (Feb. 24, 2023), https://www.axios.com/2023/02/24/chatgpt-generative-ai-legal-minefield.

[37] Antonio Pequeño IV, *George R.R. Martin and Other Big-Name Authors Sue OpenAI for Copyright Infringement*, FORBES (Sept. 20, 2023, 4:21 PM), https://www.forbes.com/sites/antoniopequenoiv/2023/09/20/george-rr-martin-and-other-big-name-authors-sue-openai-for-copyright-infringement/?sh=6ba10ed93ef8; Matt O'Brien, *Sarah Silverman and Novelists Sue ChatGPT-maker OpenAI for Ingesting Their Books*, ASSOCIATED PRESS (July 12, 2023, 2:56 PM), https://apnews.com/article/sarah-silverman-suing-chatgpt-openai-ai-8927025139a8151e26053249d1aeec20**;** James Vincent, *Getty Images Sues AI Art Generator Stable Diffusion in the US for Copyright Infringement*, VERGE (Feb. 6, 2023, 11:56 AM), https://www.theverge.com/2023/2/6/23587393/ai-art-copyright-lawsuit-getty-images-stable-diffusion.

[38] 17 U.S.C. § 106.

[39] 17 U.S.C. § 501.

[40] Sturdza v. United Arab Emirates, 281 F.3d 1287, 1295 (D.C. Cir. 2002).

[41] Narell v. Freeman, 872 F.2d 907, 910 (9th Cir. 1989) (citing Baxter v. MCA Inc., 812 F.2d 421, 423 (9th Cir. 1987)); *Sturdza*, 281 F.3d at 1295.

[42] Christopher T. Zirpoli, *Generative Artificial Intelligence and Copyright Law*, CONG. RSCH. SERV. 4, https://crsreports.congress.gov/product/pdf/LSB/LSB10922#:~:text=AI%20programs%20might%20also%20infringe,created%20 "substantially%20similar"%20outputs (last updated Sept. 29, 2023).

perspective and consider whether the infringing work replicates the "total concept and feel of the work."[43] They may also put themselves in the perspective of an "ordinary, reasonable person" and determine if such a person would presume the accused infringer copied the original work.[44] Additionally, courts will "consider the qualitative and quantitative significance of the copied portion in relation to the plaintiff's work as a whole."[45]

An owner's exclusive right to their copyrighted works, however, is not absolute and is subject to limitations.[46] First, the copyright statute permits copies "for purposes such as criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research".[47] These purposes, however, do not automatically constitute fair use and are limited by additional factors such as whether the copier is profiting from the use.[48] The statute limits protection to "works of authorship" and specific subject matter within these works.[49] Additionally, not every component of a created work is necessarily copyrightable. Certain elements may be barred from protection if they are an "idea, procedure, process, system, method of operation, concept, principle, or discovery, regardless of the form in which it is described, explained, illustrated, or embodied in such work".[50] Thus, even if a work is copyrightable as a whole, certain parts of it may be freely replicated by non-copyright holders. In art forms where the creative elements may be more nuanced, such as photography, the copyrightable parts of the work may be an amalgamation of these elements rather than the subject of the work itself. This was established in the early days of photography when a court found that a photograph of a famous person was copyrightable:

[the photographer captured the image] entirely from his own original mental conception, to which he gave visible form by posing the said Oscar Wilde in front of the camera, selecting and arranging the costume, draperies, and other various accessories in said photograph, arranging the subject so as to present graceful outlines, arranging and disposing the light and shade, suggesting and

[43] Cavalier v. Random House, Inc., 297 F.3d 815, 827 (9th Cir. 2002).

[44] Griner v. King, 568 F. Supp. 3d 978, 994–95 (N.D. Iowa 2021).

[45] Monbo v. Nathan, 623 F. Supp. 3d 56, 89 (E.D.N.Y. 2022) (citing Newton v. Diamond, 388 F.3d 1189, 1195 (9th Cir. 2004)).

[46] 17 U.S.C. § 107.

[47] *Id.*

[48] Princeton Univ. Press v. Mich. Document Servs., 99 F.3d 1381, 1391 (6th Cir. 1996) ("Professor-selected excerpts from scholarly texts copied into coursepacks sold to students on per page basis for use in courses taught by professors making selections does not constitute fair use . . . .").

[49] 17 U.S.C. § 102.

[50] 17 U.S.C. § 102(b).

evoking the desired expression, and from such disposition, arrangement, or representation, made entirely by plaintiff, he produced the picture in suit.[51]

Second, the fair use doctrine limits the exclusive rights copyright owners have over their works and permits certain uses of those works by non-owners.[52] Courts consider four factors when determining whether a non-owner's use is permittable under the fair use doctrine. These are the purpose and character of the use, the nature of the copyrighted work, the amount of the copyrighted work used, and the effect of the use on the market from which the copyrighted work originates.[53] Yet, the fair use doctrine is not impervious to deeper complexities and nuanced applications. In *Google LLC v. Oracle America, Inc.*, the Supreme Court stated that the application and ultimate success of a fair use defense can be significantly influenced by the individual facts surrounding each case.[54] A defendant will prevail in a copyright infringement action if the owner does not show copying took place or the copying is otherwise permitted under a defense like the fair use doctrine.[55]

Purpose and character refer to how the non-owner is using the copies.[56] Educational, non-profit copies are favored for fair use over for-profit works that merely replicate the original work.[57] Whether and to what degree a copy transforms the original work stands as a significant yet complex subfactor within a purpose and character analysis.[58] The nature of the work depends on whether the work is comprised of factual or creative elements.[59] Facts and the non-creative presentation of facts are not copyrightable.[60] In reviewing the amount used, courts will look at the number of copied elements or the quality of the copied elements in relation to the copyrighted work as a whole.[61] The market impact assessment may draw upon several economic considerations, but courts focus on "the effect of the [non-owner's] use upon the potential market for or value of the copyrighted work."[62] A non-owner copying in accordance with one

[51] Burrow-Giles Lithographic Co. v. Sarony, 111 U.S. 53, 60 (1884).
[52] 17 U.S.C. § 107.
[53] *Id.*
[54] Google LLC v. Oracle Am., Inc., 141 S. Ct. 1183, 1197 (2021) [hereinafter *Google*] ("[C]ourts must apply [the fair use doctrine] in light of the sometimes conflicting aims of copyright law, and that its application may well vary depending upon context.").
[55] 4 MELVILLE B. NIMMER & DAVID NIMMER, *Nimmer on Copyright* § 13.04 (rev. ed. 2024).
[56] *Copyright and Scholarship: Fair Use*, B.C. LIBR., https://libguides.bc.edu/copyright/fairuse (last updated May 9, 2023, 2:25 PM).
[57] *Id.*
[58] NIMMER & NIMMER, *supra* note 55, § 13F.05.
[59] *Copyright and Scholarship: Fair Use*, *supra* note 56.
[60] *See* Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 362–63 (1991) (explaining that a telephone company compiling information alphabetically in a phone book was not protectable and competitor's use did not constitute copyright infringement).
[61] *Copyright and Scholarship: Fair Use*, *supra* note 56.
[62] *Id.*; 17 U.S.C. § 107(4).

of these factors does not automatically constitute a fair use.[63] All four factors are weighed against one another in consideration of the unique fact scenario.[64]

D. AI'S CURRENT POSITION IN THE REALM OF COPYRIGHT

Although AI's complications are a relatively recent issue, authoritative bodies have made some decisions regarding AI's stance in copyright law. Courts have already started to mark out the scope of AI's capabilities to produce copyrightable material.[65] In a case in which an AI developer tried to register a work that was "autonomously created by a computer algorithm running on a machine" for copyright protection, the court upheld the Register's rejection of the application.[66] The court found that an AI system is not able to hold a copyright as an entity in its own right.[67] In this decision, the United States Patent and Trademark Office (USPTO) reaffirmed the long-held notion that only humans with human-made works can hold a copyright in the United States.[68] In light of this, however, courts and the U.S. Copyright Office have left open the possibility for creators to get copyrights for AI-assisted works in the future.[69] The key consideration for the USPTO in these cases is the quantity of human creativity added to AI-generated work. At least currently, the USPTO does not appreciate AI output as essentially equal to products of human ingenuity to warrant a copyright without human contribution.[70] Whether this notion will be a factor in assessing AI's ability to infringe on copyrighted works is yet to be determined.

The question of whether AI infringes on the copyrights of its training material is much murkier, but some courts have begun to outline the scope of

[63] *Copyright and Scholarship: Fair Use*, *supra* note 56.

[64] *Id.*

[65] *See, e.g.*, Thaler v. Perlmutter, No. 22-1564 (BAH), 2023 U.S. Dist. LEXIS 145823 (D.D.C. Aug. 18, 2023).

[66] *Id.* at *2–3.

[67] *Id.* at *20.

[68] *Id.* at *20; *see* Naruto v. Slater, 888 F.3d 418, 420 (9th Cir. 2018) (stating all animals cannot sue for copyright infringement because they are not human and therefore are not protected under Copyright Act).

[69] *Perlmutter*, 2023 U.S. Dist. LEXIS at *19–20 ("[T]his case presents *only* the question of whether a work generated autonomously by a computer system is eligible for copyright. *In the absence of any human involvement in the creation of the work*, the clear and straightforward answer is . . . [n]o.") (emphasis added); Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence, 88 Fed. Reg. 16190, 16192 (Mar. 16, 2023) ("Based on the Office's understanding of the generative AI technologies currently available, users do not exercise ultimate creative control over how such systems interpret prompts and generate material.").

[70] *Perlmutter*, 2023 U.S. Dist. LEXIS at *20.

infringement claims.[71] Naturally, if facts cannot be copyrighted then an AI simply restating a fact it gained from its training material does not infringe on a copyright.[72] The principal issue lies within the arguments presented by both AI owners and copyright owners. The USPTO recognizes the central argument between AI owners and copyright holders of the AI training material:

The ingestion of copyrighted works for purposes of machine learning will almost by definition involve the reproduction of entire works or substantial portions thereof. Accordingly, whether this constitutes copyright infringement will generally be determined by considering the applicability of the fair use doctrine, an exception set forth in section 107 of the Copyright Act, 17 U.S.C. § 107.[73]

Copyright owners argue that AI programmers are making copies of the owners' protected work in violation of their exclusive rights.[74] Under the USPTO's assessment, many generative AI companies are raising a fair use defense for using copyrighted works to train their models.[75] Fair use is an affirmative defense,[76] so AI companies have the burden to show their use of copyrighted materials weighs in their favor.[77]

In consideration of the four factors for a fair use defense, AI owners must demonstrate that their use of the copyrighted work weighs in favor of fair use.[78] Whether there is a protectable substantial likeness between the training material and the AI's output will likely weigh heavily in courts reviewing AI's fair use defense.[79] One judge has already threatened to dismiss a class action claim

[71] *See* Naruto v. Slater, 888 F.3d 418, 420 (9th Cir. 2018) (stating all animals cannot sue for copyright infringement because they are not human and therefore are not protected under Copyright Act).

[72] *See* Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 362–63 (1991) (explaining that a telephone company compiling information alphabetically in a phone book was not protectable and competitor's use did not constitute copyright infringement)

[73] *Public Views on Artificial Intelligence and Intellectual Property Policy*, USPTO 24 (Oct. 2020), https://www.uspto.gov/sites/default/files/documents/USPTO_AI-Report_2020-10-07.pdf.

[74] Pequeño IV, *supra* note 37.

[75] Amended Answer at 39–40, Thomson Reuters Enter. v. Ross Intel., Inc., 1:20-cv-00613-LPS (D. Del. filed Jan. 25, 2021); Bryson Masse, *OpenAI seeks to Dismiss Majority of Sarah Silverman's and Author's Claims in ChatGPT Lawsuits*, VENTUREBEAT (Aug. 29, 2023, 12:45 PM), https://venturebeat.com/ai/openai-seeks-to-dismiss-majority-of-sarah-silvermans-and-authors-claims-in-chatgpt-lawsuits/.

[76] 17 U.S.C. § 107.

[77] *Public Views on Artificial Intelligence and Intellectual Property Policy*, *supra* note 73.

[78] *Id.*

[79] Blake Brittain, *US Judge Finds Flaws in Artists' Lawsuit Against AI Companies*, REUTERS (July 19, 2023, 7:32 PM), https://www.reuters.com/legal/litigation/us-judge-finds-flaws-artists-lawsuit-against-ai-companies-2023-07-19/ ("I don't think the claim regarding output images is plausible at the moment, because there's no substantial similarity' between images created by the artists and the AI systems, [Judge] Orrick said.").

because the copyright owners did not initially show a similarity between their work and the AI's output.[80] If AI companies present facts that support a finding of fair use, a jury will decide whether the AI's copying is fair use or not.[81]

E. WHAT IS AT STAKE

Copyright owners and their supporters urge courts to find against AI companies for using their copyrighted works without first gaining a license from the owners.[82] Without licenses, AI owners would be limited to training their models on free use or public domain material. Though the public domain is a vast collection of formerly copyrighted works that grows annually, the works entering the public domain are at least seventy years old.[83] Proponents of the fair use defense for AI training argue that limiting AIs to train on only non-copyrighted material will severely stifle AI's capabilities as a tool.[84] This concern, while valid, should not precede the legal rights of copyright owners. In the instances of AI companies freely copying protected work to produce for-profit outputs, the companies should be required to obtain permission from the owners of the protected works.

III. ANALYSIS

The federal government should uphold the exclusive rights of copyright holders and restrain AI owners from using protected material to train their AIs. AI companies' concerns over stalling productivity, while valid, should not precede the legal rights of copyright owners. Rather, the current requirement of gaining permission before using copyrighted works should apply to use in AI technology as it does in any other medium. AI companies should not be allowed to profit from the derivatives of copyrighted works simply because those works were only used as intermediate copies. The process of training an AI should be considered an infringement of the copyrights of protected works when the training material is apparent in the output and the output lacks a transformative quality. A licensing requirement will preserve both the rights of copyright owners and the expandability of AI as a tool for innovation.

---

[80] *Id.*

[81] Memorandum Op. at 23, Thomson Reuters Enter. v. Ross Intel., Inc., 1:20-cv-00613-SB (D. Del. filed Sept. 25, 2023).

[82] Pequeño IV, *supra* note 37.

[83] *How Long Does a Copyright Last*, Copyright Alliance, https://copyrightalliance.org/faqs/how-long-does-copyright-last/ (last visited Sept. 15, 2024).

[84] Bloomberg Law, *ChatGPT and Generative AI Are Hits! Can Copyright Law Stop Them?*, YOUTUBE (June 26, 2023), https://www.youtube.com/watch?v=bRqwTP2eKJY.

A. THE TUMULTUOUS OUTLOOK FOR GENERATIVE AI FAIR USE

Courts will likely find that AI's use of copyrighted works as training material is not immediately exempt from copyright laws. Hany Farid, a computer science professor, states that "[training material] serve[s] as a kind of background instruction that allows the model to infer concepts like color, objects and artistic style."[85] These concepts, however, are not protected on their own.[86] No one owns the concept of a clock and may therefore bar anyone else from making an image of a clock, nor does anyone own the color red. Anyone may study the Surrealist movement in the art world and recreate the movement's style of bizarre imagery and nonsensical scenes.[87] Yet, no one may copy Salvador Dalí's *The Persistence of Memory* painting or any of its protected elements without first obtaining a license.[88] Likewise, an AI should not be permitted to copy protected works in their outputs. A fair use analysis will likely favor copyright owners, specifically in factors one and four weighing considerably against AI companies.

1. Purpose and Character: The Transformative Requirement.

Certain generative AI companies do not use copies of protected works "for purposes such as criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research[.]"[89] Even if teaching is stretched to include training AI computer systems to recognize patterns, for-profit companies should be subject to the limitations against commercializing copies. One court has already determined that for-profit AI companies cannot rely on an educational, non-commercial defense and must show the resulting AI outputs are sufficiently transformative.[90] Therefore, this element weighs against AI companies that require paid subscriptions to access their models, because this particular use is commercial and not educational.

OpenAI, one of the several AI companies with pending infringement suits, claims that AI's use of training material is "highly transformative[.]"[91] It makes

[85] Isaacs-Thomas, *supra* note 15, at 2.

[86] 17 U.S.C. § 102(b).

[87] *See generally* Julie Lagier, *The New Surrealist Artists*, ARTSPER MAGAZINE (Mar. 20, 2022), https://blog.artsper.com/en/get-inspired/contemporary-surrealism/ (showcasing recent artwork heavily inspired by the Surrealist movement).

[88] *Reproduction Rights*, FUNDACIÓ GALA–SALVADOR DALÍ, https://www.salvador-dali.org/en/services/rights-and-images/gestio-de-drets/ (last visited Oct. 29, 2024).

[89] 17 U.S.C. § 107 (stating the circumstances for which a work may be copied and be considered fair use).

[90] Thomson Reuters Enter. v. Ross Intel. Inc., No. 1:20-cv-613-SB, 2023 WL 6210901 at *5, *16 (D. Del. Sept. 25, 2023).

[91] OpenAI, LP, Comment Regarding Request for Comments on Intellectual Property Protection for Artificial Intelligence Innovation, USPTO 5, https://www.uspto.gov/sites/default/files/documents/OpenAI_RFC-84-FR-58141.pdf (citing Campbell v. Acuff-Rose Music, 510 U.S. 569, 579 (1994) (citations omitted)).

this claim based on the premise that the copies are not expressive compared to the material's original use.[92] This largely hinges on the fact that the copies are made to teach an AI system, whereas the original works were made for human entertainment.[93] OpenAI primarily relies upon the Supreme Court's opinion in *Campbell v. Acuff-Rose Music*.[94] In that case, the Court defines transformative use as a use that brings a "new expression, meaning, or message" to the original work.[95] In May of 2023, however, the Supreme Court clarified this argument in relation to copyrighted works:

Campbell cannot be read to mean that §107(1) weighs in favor of any use that adds new expression, meaning, or message. Otherwise, "transformative use" would swallow the copyright owner's exclusive right to prepare derivative works, as many derivative works that "recast, transfor[m] or adap[t]" the original, §101, add a new expression of some kind. The meaning of a secondary work, as reasonably can be perceived, should be considered to the extent necessary to determine whether the purpose of the use is distinct from the original.[96]

Furthermore, *Campbell* was a parody case in which a rap group used the melody of the song "Pretty Woman" with their own lyrics added for most of the song.[97] Although the group copied protected elements of the original artist's song, the court viewed the use as fair under the "comment" or "criticism" purpose.[98] Although AI companies refer to the copying and utilizing of copyrighted works as a training process, courts are unlikely to view their use as a teaching purpose. As machines, AI systems do not make creative, criticizing, or otherwise expansive decisions when analyzing their training material. They merely recognize patterns within a data set and replicate those patterns in an output.[99] Using copyrighted works to train AI systems does not fall into one of protected purposes, and, thus, turns on a question of transformation.

Secondly, the lack of transformative application in generative AI ultimately distinguishes these cases from fair use judicial precedent. In *Google*, Google copied lines of Oracle's Sun Java API computer code and applied those copies to Google's smartphone program.[100] The Court determined that Google's copying of the code was fair use because the copying transformed the code's

[92] *Id.*
[93] *Id.*
[94] *Id.*
[95] Campbell v. Acuff-Rose Music, 510 U.S. 569, 579 (1994) [hereinafter *Campbell*].
[96] Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith, 598 U.S. 508, 511-12 (2023) [hereinafter *Warhol*].
[97] *Campbell*, 510 U.S. at 572-73.
[98] *See Campbell*, 510 U.S. at 583 ("While we might not assign a high rank to the parodic element here, we think it fair to say that 2 Live Crew's song reasonably could be perceived as commenting on the original or criticizing it, to some degree.").
[99] *What is Generative AI?*, *supra* note 13.
[100] *See Google*, 141 U.S. at 1203.

original purpose as computer code into a new, creativity-promoting program for smartphones.[101] Though the code was copyrightable, Google's use of the copies was transformative enough to be considered protected as promoting progress under the Constitution.[102] Whether a use is transformative or not is essential to this conclusion. Like Google developing their smartphone program by copying another company's computer code, an effective generative AI training process requires that AI companies copy copyrighted material to produce an output. The similarities between the two processes end, however, at the lack of transformative qualities in generative AI output. Generative AI output, without further human intervention, lacks the creativity attractive to the court in *Google*.

In light of the Court's opinion in *Warhol*, the argument that generative AI's use of copyrighted material qualifies as transformative focuses too narrowly on the training process. Instead, the focus must be on the output and whether the resulting output fails to transform its training material. This has been a key factor in previous cases involving large collections of copyrighted images.[103] In these cases, companies made thumbnail-sized copies of protected images as a part of search engine output.[104] The Ninth Circuit favored search engine companies in part because the reduced size of the image transformed the original function of the image from entertainment to a source of information.[105] Unlike search engines, generative AI models do not expand the function of the copied image. If a protected element of the training material appears in an AI output, the element retains its entertainment function and does not gain any additional function to constitute transformation.

Depending on the level of creativity of the training material, the ultimate purpose of the copying must not be overlooked. Courts are hesitant to favor copiers on mere technicalities of a changed purpose and have recognized that a copy will need to have "sharply different objectives" to the original.[106] Therefore, even if a protected part of a work were replicated in an AI output differently than how it originally appeared, this change would not satisfy the

[101] *Id.*

[102] *Id.*

[103] *See* Perfect 10 v. Amazon.com, Inc., 508 F.3d 1146, 1165 (9th Cir. 2007) *and* Kelly v. Arriba Soft Corp., 336 F.3d 811, 818-19 (9th Cir. 2003).

[104] *Perfect 10*, 508 F.3d at 1155; *Kelly*, 336 F.3d at 815.

[105] *Perfect 10*, 508 F.3d at 1165 ("Although an image may have been created originally to serve an entertainment, aesthetic, or informative function, a search engine transforms the image into a pointer directing a user to a source of information."); *Kelly*, 336 F.3d at 819 ("[The search engine's] use of the images serves a different function than the [photographer's] use—improving access to information on the internet versus artistic expression.").

[106] Blanch v. Koons, 467 F.3d 244, 252 (2d Cir. 2006) ("The sharply different objectives that Koons had in using, and Blanch had in creating, 'Silk Sandals' confirms the transformative nature of the use."); *see also* Bill Graham Archives v. Dorling Kindersley Ltd., 448 F.3d 605, 609 (2d Cir. 2006) (referring to the change-in-purpose threshold as "plainly different from the original purpose").

sharply different threshold. Absent a parodic or otherwise transformative purpose, the purpose of the copies for AI training is not distinct from the purpose of the original works. AI companies will need to ensure the training materials' protected elements are not replicated in their AI's output. Therefore, the purpose and character factors will disfavor AI companies whose systems effectively present the work of others.

2. Nature of the Copyrighted Works

For the issue of AI using copyrighted works, the nature of the copyrighted works factor could require a tedious analysis. This question largely focuses on whether each work used as training material is comprised of more factual or more creative elements.[107] The more factual the work is the more likely a non-copyright holder will prevail on this factor of fair use.[108] Additionally, like the other factors of fair use, it applies to each instance of copying.[109] Under the extensive training process, one website could be the source of 10,000 to over 100,000 distinct images for an AI system.[110] Through this process, a single AI system could copy billions of individual works.[111] In any instance of copyright violation claims, a court would then have to apply a factual analysis to the nature of each one of the copies that appear in the output.[112] This could prove overwhelming in multiple-plaintiff cases where the copyright holders are claiming infringement for multiple different images contained in one output. In certain instances, AI copying will blatantly favor copyright owners when the copier's use under factor one is indefensible, such as non-parodic derivative works of the visual arts and creative writing.[113] Works that have a more factual nature, however, will call into question whether the AI company copied the copyrightable elements.

Moreover, the nature of AI's use of copyrighted works diverges from previous fair use cases involving similar forms of copying. Intermediate copies refer to the multiple copies one entity makes of another's work when generating a new product.[114] Although, in the past, intermediate copying has been litigated

[107] *Copyright and Scholarship: Fair Use*, *supra* note 56.

[108] *Id.*

[109] 17 U.S.C. § 107.

[110] 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/.

[111] *Id.*

[112] *Copyright and Scholarship: Fair Use*, *supra* note 56 ("The copyright statute states that the following four factors *must* be evaluated to determine . . . whether a use is fair.") (emphasis added).

[113] NIMMER & NIMMER, *supra* note 55, § 13F.06.

[114] Daniel A. Schnapp & Alexis P. Grilli, *Intermediate Copying, Artificial Intelligence and Best Practices for Counseling Music Clients in an Evolving Legal and Technological Landscape*, BLOOMBERG L. (Apr. 9, 2021), https://www.bloomberglaw.com/document/XD0QS5D0000000?jcsearch=glm45e

around computer code, it may also be involved in music production.[115] This form of copying is distinct from direct copying because the material copied within the intermediate copies is not evident in the new product.[116] As a result, courts have found intermediate copying to be defendable under the fair use doctrine.[117] Intermediate copying, however, is not *prima facie* permissible under Section 107's narrow category for permissible copying, therefore it must be analyzed under the fair use defense.[118] For intermediate copying to succeed under a fair use defense, the court must still favor the copying entity overall under the four factors.[119] In *Chapman v. Maraj*, singer Tracy Chapman sued rapper Niki Minaj for using part of Chapman's song in Minaj's demo song.[120] The court held that Minaj's demo, which did contain a copied version of Chapman's lyrics, was fair use because Minaj formed the demo as part of a non-commercialized creative process.[121] Though Minaj succeeded in her defense overall, the court found that factor two unequivocally favored Chapman due to the fact that the copied material was song lyrics.[122] Unlike computer code, which contains protected and unprotected parts that lend greater opportunities for permitted copying, "a musical composition . . . is the type of work that is at the core of Copyright's protective purpose."[123]

gdifi#jcite ("Creation of a new copyrightable work in areas such as music, artwork, and computer programming often involves the necessary step of utilizing a portion of an existing protected work. The existing work is sampled or used in an 'intermediate' capacity while the creator is working toward development of a final, new work.").

[115] *Id.*

[116] Robert V. Donahoe, *Does Intermediate Copying of Computer Software for the Purpose of Reverse Engineering a Non-Infringing Product Infringe the Copyright in the Software?*, B.C. INTELL. PROP. & TECH. F. 1 (2001), https://lira.bc.edu/files/pdf?fileid=f06b92ba-44d9-4214-bfaf-d7bbce319f92 (referencing Sega Enters. Ltd. v. Accolade, Inc., 977 F.2d 1510, 1518 (9th Cir. 1992)).

[117] *See* Sega Enters. Ltd. v. Accolade, Inc., 977 F.2d 1510, 1526 (9th Cir. 1992) (finding in favor of a company that copied another company's computer codes under all four fair use factors); *see also* Sony Comput. Ent., Inc. v. Connectix Corp., 203 F.3d 596, 602 (9th Cir. 2000) ("We conclude that, under the facts of this case and our precedent, Connectix's intermediate copying and use of Sony's copyrighted BIOS was a fair use for the purpose of gaining access to the unprotected elements of Sony's software.").

[118] *Sega Enters. Ltd.*, 977 F.2d at 1519 (holding "that intermediate copying of computer object code may infringe the exclusive rights granted to the copyright owner in Section 106 of the Copyright Act regardless of whether the end product of the copying also infringes those rights.").

[119] *Id.* at 1521 ("[The Fair Use defense] provides that particular instances of copying that otherwise would be actionable are lawful, and sets forth the factors to be considered [by the court] in determining whether the defense applies.").

[120] Chapman v. Maraj, No. 2:18-cv-09088-VAP-SSx, 2020 U.S. Dist. LEXIS 198684, at *16 (C.D. Cal. Sept. 16, 2020).

[121] *Id.* at *33.

[122] *Id.* at *30-31.

[123] *See Sega Enters. Ltd.*, 977 F.2d at 1524 (9th Cir. 1992); *Maraj, Inc.*, 2020 U.S. Dist. LEXIS, at *31.

Like Minaj's copy of Chapman's lyrics, the generative AI training process copies highly creative works that are central to the purposes of copyright protection.

3. Amount and Substantiality

Factor three will likely weigh against AI companies because the training process uses a substantial amount of the work that does not result in sufficiently transformative output. Depending on the facts of the case, non-copyright holders may be able to copy a larger amount of work to achieve a transformative or alternate purpose.[124] Although AI companies argue that the entirety of the works must be used to effectively train their systems, their use of the copyrighted material is not transformative. In other art techniques, such as collages, artists copy and use large sections of other works; courts have permitted such uses so long as the amount used in the work is proportional to achieving a different purpose than the original work.[125] In the case of AI companies' use of training material, typically the entire work is used to create an output of the same purpose. Therefore, this argument should not be upheld in this circumstance.

Allowing AI to use the entirety of works simply because it is necessary would set the development of AI technology above the protection of copyright. Neither the Copyright Act nor case law permits copying protected works simply because the copier finds it absolutely necessary to further their own purposes. In fact, it would undermine the rationale for setting aside copyright protections and effectively "eviscerate the reproduction right by requiring a distribution to take place for a violation of the reproduction right to occur."[126]

4. Effect on Market

Factor four will likely weigh the most significantly in favor of copyright owners on two points: market substitution and derivative works. Courts recognize factor four as the natural consequence of significant use under factor three:

[t]he obvious reason for this lies in the relationship between the third and the fourth factors. The larger the amount, or the more important the part, of the original that is copied, the greater the likelihood that the secondary work might serve as an effectively competing substitute for the original and might therefore diminish the original rights holder's sales and profits.[127]

---

[124] Reyes v. Wyeth Pharm., Inc., 603 F. Supp. 2d 289, 299 (D.P.R. 2009) ("[D]efendants used [the image of a sculpture] as an example of a creative work which in turn allowed them to communicate their overall message . . . . Defendants' message certainly could have been communicated in ways that did not involve the [sculpture], but . . . their message could not have been communicated as effectively by obscuring or including only part of the sculpture.").

[125] *See* Blanch v. Koons, 467 F.3d 244, 258 (2d Cir. 2006) ("[T]he amount and substantiality of Koons's copying was 'reasonable in relation to the purpose of the copying.'") (citing Campbell v. Acuff-Rose Music, 510 U.S. 569, 586 (1994)).

[126] Cala Coffman, *Does the Use of Copyrighted Works to Train AI Qualify as a Fair Use?*, COPYRIGHT ALL. (Apr. 11, 2023), https://copyrightalliance.org/copyrighted-works-training-ai-fair-use/.

[127] Authors Guild v. Google, Inc., 804 F.3d 202, 221 (2nd Cir. 2015).

In this case, the court found that Google using snippets of copyrighted books to compile their search feature was fair use because the snippets were small and did not effectively replace the books on the market.[128] In contrast, generative AI uses the entirety of the training material to generate an output that competes within the same market as the training material. For example, a user can search the stock image website for "puppies in a basket" and get roughly 1,200 photographs matching the search.[129] If an AI company were to use these photographs to train their image-generating model, then their output images would recreate the same elements. Customers of those sites would likely be interested in generative AI outputs as they contain the same qualities that attracted them to the stock image sites. Therefore, generative AI systems compete against the same markets from which they gathered the copyrighted training material.

While AI-generated works competing against human-made work does not alone affect a factor four analysis, the court does not allow the non-copyright holder to illegally overtake the copyright owners' positions in the market.[130] For example, any one stock image site competes against other similar sites. These sites, however, would not be able to legally copy images from one another to supplement their own photo collections. Additionally, they would not be able to recreate the copyrightable elements of a competitor's image for their own commercial gain.[131] AI likewise cannot do the same, but this is precisely the AI training process's goal. After successfully copying and analyzing its training material, generative AI models produce an amalgamation of the similarities among all the original works used in the training process. Thus, a user requesting an image of "puppies in baskets" would get an image that encapsulates the styles, poses, lighting, and other copyrightable elements from the training material. This concept applies to all copyrightable works. Because the AI system recreates the similarities amongst its training material and finely tunes the results to match the user input, a user could simply use the AI model to find highly specific results. If not regulated, generative AI models could produce large amounts of outputs that compete with the works used to train the model and overtake the market from which those works originate.

---

[128] *Id.* at 224 ("[W]e conclude that the snippet function does not give searchers access to effectively competing substitutes.").

[129] GETTYIMAGES, https://www.gettyimages.com/search/2/image?license=rf&alloweduse=availableforalluses&family=creative&mediatype=photography&phrase=puppies%20in%20a%20basket&sort=mostpopular (last visited Oct. 18, 2023).

[130] Campbell v. Acuff-Rose Music, 510 U.S. 569, 593 (1994).

[131] *See* Mannion v. Coors Brewing Co., 377 F. Supp. 2d 444, 463 (S.D.N.Y. 2005) ("The defendants . . . appear to have recreated much of the subject. . . then, through imitation of angle and lighting, rendered it in a similar way. The similarities . . . relate to the [Plaintiff's] originality in the rendition and the creation of the subject and therefore to its protected elements.").

Specifically, AI companies threaten to force copyright owners out of the market for derivatives of the copyright owners' works. The United States Copyright Office defines derivative works as "a work based on or derived from one or more already existing works" and explains that copyright protection extends to the new material added to the original works.[132] The right to make derivatives of copyrighted works is afforded to owners of the original work.[133] Furthermore, in *Campbell*, the court explains that "[t]he market for potential derivative uses includes only those that creators of original works would in general develop or license others to develop."[134] Open AI argues that their AI system's outputs do not infringe on the training material because evidence of the training material is not visible in the AI output.[135] This fact, however true or not, should not change the outcome of the analysis. Even if individual training material is indiscernible within the resulting output, this does not negate that the copying was an infringement in the first place.

Furthermore, generative AI learning through training material cannot be protected by likening it to the human creative process because AI algorithms cannot be creative like humans. Creativity is defined as "the ability to produce or use original and unusual ideas."[136] AI systems merely recognize statistical patterns within a large amount of data and replicate those patterns in their output. Essentially, "AI's creative processes is that the current computational creativity is systematic, not impulsive, as its human counterpart can often be. It is programmed to process information in a certain way to achieve particular results predictably, albeit in often unexpected ways."[137] AI can rearrange the elements of all the images from its training process and produce a satisfactory output. These results, however, are entirely dependent on the copying and recreation of copyright-protected elements. Therefore, AI competition within the markets of its training material is an unfair exploitation of copyrighted materials that disfavors AI companies in a factor four analysis.

When AI systems replicate elements of copyrighted material to produce competing outputs, they violate copyright owners' exclusive rights to make derivatives of their work and risk pushing human creators out of their markets. The AI's process of analyzing copyrighted works and applying its training to output production captures the process of creating derivative works. A user

---

[132] *Copyright in Derivative Works and Compilations*, U.S. Copyright Off. 1–2 (July 2020), https://www.copyright.gov/circs/circ14.pdf.

[133] 17 U.S.C. § 106.

[134] *Campbell*, 510 U.S. at 592.

[135] OpenAI, LP, *supra* note 91.

[136] *Creativity*, CAMBRIDGE ADVANCED LEARNER'S DICTIONARY & THESAURUS, https://dictionary.cambridge.org/us/dictionary/english/creativity.

[137] Chloe Preece & Hafize Çelik, *AI is a Powerful Tool, but it's Not a Replacement for Human Creativity*, WORLD ECON. F. (June 16, 2023), https://www.weforum.org/agenda/2023/06/ai-cannot-replace-human-creativity/.

requesting an AI-generated image would get an output of this request completely derived from every individual work comprising the relevant training material. By using copyrighted work without the owners' permission to train their systems, AI companies are avoiding significant development costs at the risk of taking over the very markets from which they glean their training material. Without any restrictions, this process of appropriating copyrighted elements of works threatens to deprive human authors of control over their own works and, in many cases, profit from those works. Licenses strike a balance between the rights of the human creators and the desires of the public to use creative works by ensuring creators are paid for the use of their works on terms they may set. Therefore, the law should require AI companies to obtain a license to use copyrighted works from each copyright owner whose work appears in the AI output.

B. A NEW FRONTIER: LICENSING FOR AI USE

A licensing requirement offers a mutually beneficial middle-ground between AI companies and owners of copyrighted material. While licenses may appear to be overly restrictive on AI companies, the right licensing configuration can protect the copyright while incentivizing owners to relinquish some of their rights to the progress and development of AI programs. Additionally, if AI companies adopt a licensing policy, it may reduce the amount of infringement litigation AI companies may face in the future. Presently, there are various kinds of artistic licensing agreements and options in use throughout all modes of artistic expression.[138] AI companies' desire to use copyrighted works to generate new material, especially on a large scale, presents several challenges to current licensing options.

1. The Public Domain and Creative Commons License

Currently, AI companies are free to use any material within the public domain and any material that explicitly allows for the kind of uses AI training entails.[139] Each of these categories permits the free use of a vast amount of individual works across an expansive variety of creative expressions. In 2023, the public

[138] *See* Lee Down, *Unlock Art's Potential: A Visual Artist's Guide to Licensing*, ARTS ARTISTS ARTWORK (May 16, 2023), https://artsartistsartwork.com/unlock-arts-potential-a-visual-artists-guide-to-licensing/#aioseo-a-definition-of-licensing; *Theatrical Licensing Companies*, FLOP MUSICAL, https://flopthemusical.com/theatrical-licensing-companies/ (last visited Oct. 28, 2024) ("[Theatre Rights WorldWide] exclusively grants live stage production rights to the plays and musicals in our catalogue . . . [including] the authorized performance script and music materials required to rehearse and present [TRW licensed] shows . . . .").

[139] *What Is the Public Domain?*, COPYRIGHTLAWS.COM (Mar. 7, 2023), https://www.copyrightlaws.com/what-is-the-public-domain/ ("Works that are in the public domain may be used freely, without obtaining permission from or compensating the copyright owner.").

domain alone gained thousands of books from culturally and historically significant authors.[140] AI companies, however, likely cannot feasibly rely upon the public domain and specific licensing to effectively develop generative AI systems.

The public domain offers a constantly growing amount of material, but these works are not likely to fill an AI company's need for training material on their own. Works previously protected by copyrights enter the public domain yearly, but they are largely constrained by their publication date.[141] Non-federal materials and works fixed in tangible mediums published after January 1st, 1978 automatically enter the public domain once seventy years have elapsed since the original author's death.[142] If, however, AI companies were limited to just these materials, then AI output would be entirely derived from old material. While certain more recent works are also publicly available, AI companies would still be developing their technology on grossly outdated material.[143] Thus, instead of returning the most appropriate outputs to a modern user's request in 2024, an AI system would be largely constrained to providing a matching output based on cultural and artistic expressions from before 1928. Additionally, when certain works enter the public domain their later-published derivatives may remain protected.[144] AI companies would then have to comb through the elements of each public domain work to ensure there are no lingering copyright protections before using them to train their systems. Furthermore, while some works may be labeled with the copyright symbol or another protective indicator, all fixed works are automatically protected upon creation regardless of their registration status.[145] Thus, AI companies would have to look up every image they use before adding it to their system's training-image compilation.

A similar feasibility problem arises within Creative Commons licensing. These licenses are self-applied by the owners of the copyrighted works, detailing

[140] Jennifer Jenkins, *January 1, 2023 is Public Domain Day: Works from 1927 Are Open to All!*, DUKE'S CENTER FOR STUDY PUB. DOMAIN (2022), https://web.law.duke.edu/cspd/publicdomainday/2023/.

[141] *What Is the Public Domain?*, *supra* note 139.

[142] *What is Copyright?*, COPYRIGHT.GOV., https://www.copyright.gov/what-is-copyright/#:~:text=What%20is%20copyright%20registration%3F,step%20is%20registering%20the%20work (last visited Sept. 18, 2024).

[143] *What Is the Public Domain?*, *supra* note 139 ("[P]rior to [March 1st, 1989], notice of copyright was necessary on all published works. Without this notice, the work went into the public domain.").

[144] *Id.* ("For example, Shakespeare's 'Romeo and Juliet' may be in the public domain, but a new version with annotations or illustrations may have copyright protection in these new parts of the work. However, this doesn't affect the copyright status of the public domain portions of the adaptation.").

[145] *What is Copyright?*, *supra* note 142.

the exact terms of use for anyone desiring to copy the work.[146] Each version of a license may impose new duties and require new investigations and therefore is likely to increase external information costs. Within this system, AI would be limited to using only materials that allow both commercial and derivative uses. This amounts to only two of the six possible licenses in the Creative Commons.[147] Furthermore, within those materials that permit both uses, any works not dedicated to the public domain require the new user to give credit to the original creator.[148] While AI companies are capable of storing and releasing the source information of their training material, it could easily become impractical to attribute credit individually to potentially millions of original artists.[149] Additionally, it may defeat the creator's purpose in requesting attribution if their name appears as a single line amongst the sea of others whose work went into the creation of an AI output. AI companies can still utilize both public domain and Creative Commons licensed works, but they likely will not be able to utilize these works on a major scale. For that, AI companies will need to rely on mass licensing agreements that suit the specific uses entailed in AI training.

2. The Potential for Private Licenses

The current AI training method presents a relatively unique issue in that it requires copious amounts of copyrighted material to produce a derivative output. Likewise, it presents a correlatively unique problem in obtaining licenses to use these materials. Private licensing offers a solution to generative AI companies' need to amass vast amounts of permission to effectively train their programs. Additionally, it encourages human creators to invest their talents into a nascent technology. This can be done by companies and copyright owners utilizing an already existing system for large-scale licensing.

Individual licenses allow AI companies to circumvent the use and attribution issues in the Creative Commons by directly laying out the terms of use in the contract parameters. Additionally, by defining precisely how the work will be used, both parties can achieve a deal where each of their needs and desires are fulfilled. Yet, the general format of artistic licensing agreements will still need to adapt to the nature of the AI training process, especially in questions of duration, revocability, and exclusivity.

One of the most significant issues in private licenses that are complicated by the nature of generative AI is the duration of the license. As one law firm estimates, "[m]any agreements will ask for a license duration of 12 to 24 months

146 *Copyright and Intellectual Property Toolkit*, UNIV. OF PITTSBURG LIBR. SYS., https://pitt.libguides.com/copyright/licenses (last updated July 17, 2024, 10:57 AM).
147 *Id.*
148 *Id.*
149 Baio, *supra* note 110.

with the option to renew."[150] In the past, artistic licenses were attractive to artists for several economic and reputational purposes:

[Artists can] extend the reach of their creativity and expand their audience beyond the confines of the traditional art market . . . . [A]rtists can see their art showcased on a wide range of products, merchandise, and publications, thereby gaining exposure to new demographics and markets. This . . . not only enhances their reputation but also creates opportunities for collaboration with established brands, businesses, and individuals. Additionally, licensing offers artists a valuable source of income through licensing fees and royalties. It allows them to generate revenue from their art multiple times over, as their work is reproduced and utilized in various commercial and non-commercial contexts.[151]

Generative AI, however, does not use copyrighted materials in these traditional ways. Rather than showcasing an artist's work, generative AI programs learn from their training material to better correlate an output with a user's request. Once an AI system is trained on a piece of data, there is no way to "reverse the changes induced in [the AI] system by a single data point at the request of a data owner".[152] The system cannot simply stop using a data point should a license expire, or the artist wish to revoke permission. Therefore, an owner licensing their work to an AI company would be permanently relinquishing their work to the company regardless of how long the license actually lasts. AI experts, however, suggest that removing training data upon the original owner's request may not be an impossible task for programmers.[153] Although removing singular data points can protect artists from perpetually relinquishing their rights, this method may be overburdensome for AI programmers. Removing one data point may have minimal effects on a program, but if an AI company has thousands of licenses expiring intermittently, this could be difficult to constantly track and implement. Perpetual licenses resolve this issue.

Perpetual licenses grant the licensee continuous permission to use a copyrighted work until revoked.[154] Though not historically a parameter in visual

[150] Aaron Pierce, *Artist Licensing Agreements: Drawing the Line on Creative Usage*, PIERCE & KWOK LLP (Oct. 1, 2021), https://piercekwok.com/artist-licensing-agreements-drawing-the-line-on-creative-usage/.

[151] Down, *supra* note 138.

[152] *We Forgot to Give Neural Networks the Ability to Forget*, FORBES (Jan. 25, 2023, 11:30 AM), https://www.forbes.com/sites/ashoka/2023/01/25/we-forgot-to-give-neural-networks-the-ability-to-forget/?sh=2a7ff0026853.

[153] *Id.* ("There are adjacent ideas to get inspiration from such as the concept of 'catastrophic forgetting,' the tendency of AI models to forget previously learned information upon learning new information.").

[154] Ryen Rasmus, *What Do the Terms Mean in My Intellectual Property License, and Does It Protect Me?*, LIPP L. (Jan. 12, 2022), https://www.lipplawfirm.com/intellectual-property-license-terms/.

art licenses, perpetual licenses have been utilized in music and virtual books.[155] In this case, the owner of the song or book grants the purchaser the right to download a copy of the work and use that copy for personal entertainment.[156] While these particular licenses would likely not permit the commercialized or derivative uses necessary for AI training, the permanency of the agreement can reduce the costs of single-data extraction.[157] Perpetuity suits the continuous use of copyrighted material in the AI training process. Additionally, revocability clauses protect copyright owners from unauthorized abuses by acting as a kill switch to end the license under certain contractual provisions.[158] Within the freedom of contract, AI companies and copyright owners can come to an agreement with a perpetual license that satisfies both parties' objectives.[159]

Despite these advantages, AI companies are still faced with the task of implementing such a system on a large enough scale to satisfy the vast amount of material needed for the AI training process. Typically, the more permission a licensor gives, the more expensive that license will be for the licensee.[160] For AI companies seeking to amass licenses for millions of works created by thousands of different people, individually negotiating each one can be expensive for both companies and copyright owners.[161] Therefore, to mitigate these expenses, AI companies and copyright owners may benefit from third-party organizations geared toward supplying blanket licenses for AI use.

3. The Blanket License

Under a licensing requirement, a blanket licensing system provides the maximum benefits to AI companies and copyright owners. Courts have already validated the necessity of these types of licenses in similar industries:

A middleman with a blanket license was an obvious necessity [in music broadcasting] if the thousands of individual negotiations, a virtual impossibility, were to be avoided. Also, individual fees for the use of individual compositions would presuppose an intricate schedule of fees and uses, as well as a difficult and

---

[155] *Perpetual License Agreement,* CONTRACTSCOUNSEL, https://www.contractscounsel.com/t/us/perpetual-license-agreement (last visited Sept. 15, 2024).

[156] *Id.*

[157] *Copyright and Intellectual Property Toolkit, supra* note 146 (explaining that retraining an AI system after removing a data point "is not a practical solution because it requires too much computing power and it is too resource intensive.").

[158] Rasmus, *supra* note 154 ("[R]evocable licenses may be terminated by the licensor (either at the licensor's discretion or in the event of later-occurring conditions, like breach of contract or lapse of the license's term).").

[159] *See generally* Pierce, *supra* note 150.

[160] Pierce, *supra* note 150 *("[G]enerally speaking, the broader the permission, the greater the compensation.").*

[161] *Intellectual Property License Agreement Cost,* CONTRACTSCOUNSEL, https://www.contractscounsel.com/b/intellectual-property-license-agreement-cost (last visited Oct. 21, 2023).

expensive reporting problem for the user and policing task for the copyright owner.[162]

Instead of limiting permission to a single creator and their select works, blanket licenses grant a licensee permission to use an amalgamation of work retained in a single portfolio for a single fee.[163] Currently, blanket licensing is predominantly popular in the music industry and frequently used in music streaming services.[164] Like AI companies, music streaming services can be more attractive to consumers the more work they can obtain through licensing.[165] The more material an AI has available through its training process, the better it can construct an accurate output. Currently, individual hubs for potential training material have formed licensing agreements with OpenAI and the same AI company has expressed a willingness to invest in more licenses.[166] A licensing mandate will open the door for licensing intermediaries, like BMI and ASCAP, to unite the interests of generative IA companies and copyright holders.

4. The Future of AI Under Licensing Requirements

AI has great potential as a tool; like the camera, generative AI can be a tool for artistic expression but must adhere to the fairness of copyright laws. A photographer cannot replicate the creative elements of another photograph, and generative AI companies should be afforded no more rights than that. Not all generative AI outputs will infringe upon the rights of their training material. Yet, the sheer number of images and intricacies in the training process may put a strain on generative AI companies to ensure their models comply with copyright laws. As is the case for many infringement cases, the current cases against AI companies are highly fact-intensive and will vary for each AI output in question.

---

[162] Broad. Music, Inc. v. Columbia Broad. Sys., Inc., 441 U.S. 1, 20 (1979).

[163] Maralee Buttery, *Blanket Licensing: A Proposal for the Protection and Encouragement of Artistic Endeavor*, 83 COLUM. L. REV. 1245, 1246 (1983).

[164] *See About Us*, ASCAP, https://www.ascap.com/about-us (last visited Oct. 29, 2024); *About,* BMI, https://www.bmi.com/about (last visited Oct. 29, 2024); *see also Information for Artists Submitting to Pandora*, PANDORA, https://help.pandora.com/s/article/Information-for-Artists-Submitting-to-Pandora-1519949298669?language=en_US#anyone (last visited Oct. 29, 2024) (listing the digital aggregators Pandora partners with for licensing); *Intellectual Property Policy*, PANDORA, https://www.pandora.com/legal/intellectual-property/ (last visited Oct. 29, 2024) ("Most content available on Pandora is provided and licensed to Pandora from third party content providers . . . in combination with blanket, direct, and/or compulsory publishing licenses.").

[165] Antonio Villas-Boas & Sky Gould, *How the Top Streaming Services Compare*, BUS. INSIDER (Mar. 30, 2016, 5:27 PM), https://www.businessinsider.com/how-the-top-music-streaming-services-compare-2016-3.

[166] Joe Panettieri, *Generative AI Lawsuits Timeline: Legal Cases vs. OpenAI, Microsoft, Anthropic and More*, SUSTAINABLE TECH PARTNER (last updated Jan. 8, 2024), https://sustainabletechpartner.com/topics/ai/generative-ai-lawsuit-timeline/ (reporting that both Shutterstock and the Associated Press have licensed their work to OpenAI for training purposes).

Implementing licensing agreements may add an extra layer of insurance against infringing AI outputs.

IV. CONCLUSION

The courts need to preserve the exclusive rights of copyright holders. These rights are intended to protect the holders' creative outputs and ensure they can reap the commercial rewards for their efforts as they desire. Some creators will want to maintain full protection of their works while others will want to exercise their rights under licensing agreements. They should be able to prevent their work from being replicated in generative AI outputs. Authors and artists should not be pushed out of the market by a machine that effectively and necessarily feeds off their collective creativity. Being completely reliant upon copied images to produce a sensible output, AI would essentially starve itself if it were permitted to push its training material source out of the market. Therefore, for the sustainment of both human creators and AI systems, generative AI companies should have to obtain a license to use copyrighted material when those materials appear in AI outputs. Just as AI systems are becoming an everyday tool for people in many different industries, AI-human creator licenses may become essential to the expansion of both parties.