NORTH CAROLINA JOURNAL OF LAW & TECHNOLOGY
VOLUME 25, ISSUE 3: APRIL 2024

GENERATIVE AI AND COPYRIGHT LAW: A MISALIGNMENT THAT COULD LEAD TO THE PRIVATIZATION OF COPYRIGHT ENFORCEMENT

John T. Kivus*

The Article discusses the challenges and implications of applying existing copyright law to a subset of artificial intelligence ("AI") that creates new data known as generative AI ("Generative AI"). Specifically, the Article examines the mismatch between copyright law and the unique legal complexities that arise from the training and use of Generative AI. The Article argues that this mismatch could lead to an increased privatization of copyright enforcement.

Additionally, this Article discusses: (1) copyright law fundamentals crucial to understanding the law's application to Generative AI; (2) the technological underpinnings of Generative AI that differentiate it from previous new technologies; and (3) the rapidly growing body of disputes and decisions struggling to apply existing copyright law to Generative AI inputs and outputs. The Article also takes an in-depth look at how attempting to apply copyright law's fair use defense and human authorship requirement

* John is a partner at Morningstar Law Group who focuses his practice on software related issues and other intellectual property matters. His writing on software is available at http://softwarelitigation.net. He would like to thank the partners at his firm who have been understanding as he dedicated his time to preparing this Article—especially Chris Jackson. Additionally, he would like to thank the editors of the North Carolina Journal of Law and Technology for the opportunity to participate in its symposium on intellectual property law and artificial intelligence. He would also like to thank his copyright professor, Deborah Gerhardt, for providing him with his foundational copyright knowledge, and Nilay Patel, editor-in-chief at The Verge, for his common refrain that "copyright law is based on making copies," which has been a grounding principle during his legal career and is found thematically throughout this Article. Lastly, he would like to thank his wife, Elyssa, for her ongoing support and encouragement, particularly while he spent an extraordinary amount of time working to finish up this Article.

to Generative AI pushes against the established boundaries of how copyright law usually deals with emerging technologies.

The Article posits that the rapid growth and development of Generative AI requires a modernization or adaptation of copyright law to address Generative AI's unique challenges and questions. It explores the consequences of denying copyright protection to works created with Generative AI, including computer software, and warns that copyright enforcement may be left to private entities that lack the benefits and safeguards of the court system if existing copyright law is not modernized.

TABLE OF CONTENTS

I. INTRODUCTION.....449
II. THE FOUNDATIONS OF COPYRIGHT LAW 450
A. Elements of a Copyrightable Work: Originality and Human Authorship.....451
1. Works by Non-Human Creators Are Not Copyrightable 452
2. A Garden Lacks Human Authorship and Fixation..453
B. Copyright Law's Previous Adaptations to New Technologies 455
1. Photography.....455
2. Digitization of Printed Materials for Online Use...457
III. A BRIEF INTRODUCTION TO GENERATIVE AI.....460
A. Key Generative AI Concepts.....461
1. Large Language Models 461
2. Word Vectors or Word Embedding 462
3. Transformer Models.....464
B. Training Generative AI Models 465
IV. DIFFICULTIES APPLYING EXISTING COPYRIGHT LAW TO GENERATIVE AI.....467
A. Generative AI Lawsuits.....468
1. Initial Lawsuits Focus on Training AI Models 468
2. The Focus of Lawsuits Expands to Include Outputs 470
3. Generative AI Companies Rely on Fair Use.....475
B. The Copyright Office Denies Applications for Generative AI Works.....479

1. Works Created Solely by Generative AI 481
2. Portions of a Larger Work Created Solely by Generative AI 482
3. Works That Include Both Human and Generative AI Contributions 484
C. Consequences of Denying Works Created with Generative AI Copyright Protection 485
V. THE PERILOUS ROAD TO PRIVATE COPYRIGHT ENFORCEMENT 487
VI. CONCLUSION 493

I. INTRODUCTION

The concept of artificial intelligence ("AI") has obtained a foothold outside of the technology sphere that far exceeds other recent technological trends like Web3, blockchain, or cryptocurrency. This interest seems primarily concentrated on generative artificial intelligence ("Generative AI") systems, which attained high levels of notoriety in 2022 with the public release of newer versions of models such as Stable Diffusion and ChatGPT. The increased interest has resulted in the rapid development of Generative AI and the corresponding increased scrutiny.

This scrutiny applies to the inputs used to train Generative AI, as well as the outputs produced by it. Generative AI inputs and outputs implicate existing United States ("U.S.") copyright law. For instance, the unlicensed use of copyrighted materials to train the models can establish prima facie copyright infringement. Additionally, questions about fair use and the impact of Generative AI outputs on fair use analysis are being raised, but answers to these questions are unclear at this time.

Courts have historically been able to grapple with issues involving emerging technology and copyright, even when there was no law directly on point. The copyright questions raised by Generative AI, however, combined with the size and influence of the companies involved in the space, make the existing copyright law a poor fit for resolving current and potential Generative AI copyright disputes. In other words, while prior courts have been able to address

and answer questions about new technologies solely by adapting existing law, this strategy will not work for the questions presented by Generative AI. A failure to modernize copyright law to address Generative AI issues will lead to an increase in private copyright enforcement that lacks the benefits and safeguards of our court system. This trend has already begun.

This Article examines the difficulty in applying existing copyright law jurisprudence to issues regarding Generative AI and raises concerns regarding the potential for privatization of copyright infringement if copyright law is not modernized or, at the very least, adapted. The Article first discusses the foundations of copyright law in Part I and gives a high-level introduction to

Generative AI in Part II. Part III evaluates the current state of disputes and opinions involving Generative AI issues, and Part IV documents the examples of private copyright enforcement that are starting to develop in the void left by copyright law as it exists presently. Part V concludes that copyright law, as it exists presently, is not equipped to handle Generative AI.

II. THE FOUNDATIONS OF COPYRIGHT LAW

"The source of Congress' power to enact copyright laws is Article I, § 8, cl. 8 of the Constitution, which authorizes Congress to 'secur[e] for limited Times to Authors . . . the exclusive Right to their respective Writings.' "[1] Much as with patents, the economic theory behind copyrights is that the best way to encourage authors to create works beneficial to the public is to provide ways for those authors to benefit from the works personally.[2] This concept was so

[1] Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340 (1991); *see also* 1 MELVILLE B. NIMMER & DAVID NIMMER, NIMMER ON COPYRIGHT § 1.02 (2023). Though some point to Article I, § 8's mention of "useful arts" when discussing copyright, the Supreme Court clarified in *Graham v. John Deere Co*., 383 U.S. 1 (1966), that, in colonial usage, "useful arts" referred to the work of inventors, *id.* at 5 n.1 (1966) (citing RICHARD CROSBY DE WOLF, AN OUTLINE OF COPYRIGHT LAW 15 (1925) (other citations omitted)); *see also* NIMMER & NIMMER, *supra*, § 1.03.

[2] Mazer v. Stein, 347 U.S. 201, 219 (1954); *see also* NIMMER & NIMMER, *supra* note 4, § 1.03(A)(1).

important to the Founders that they enshrined it in our country's founding documents.

The current and operative Copyright Act was enacted in 1976 and became effective on January 1, 1978.[3] As defined in that Act, copyright protection extends to "original works of authorship fixed in any tangible medium of expression, now known or later developed, from which they can be perceived, reproduced, or otherwise communicated."[4] A work is deemed "fixed" under the Act "when its embodiment in a copy . . . is sufficiently permanent or stable to permit it to be perceived, reproduced, or otherwise communicated for a period of more than transitory duration."[5] Courts and the Copyright Office have applied the originality, authorship, and fixation requirements to exclude creations made with Generative AI from copyright protection.[6]

A. Elements of a Copyrightable Work: Originality and Human Authorship

As Justice Sandra Day O'Connor explained in *Feist,*

> [t]he *sine qua non* of copyright is originality. To qualify for copyright protection, a work must be original to the author. Original, as the term is used in copyright, means only that the work was independently created by the author (as opposed to copied from other works), and that it possesses at least some minimal degree of creativity.[7]

Justice O'Connor further expanded on this concept by stating that writings "are original, and are founded in the creative powers of the mind. The writings that are to be protected are the fruits of intellectual labor, embodied in the form of books, prints, engravings, and the like."[8] She based this decision, in part, on the Supreme Court's previous ruling in *Burrow-Giles Lithographic Co. v.*

---

[3] 17 U.S.C. § 101 (originally enacted as Copyright Act of 1976, Pub. L. No. 94-553, § 102, 90 Stat. 2541(1976)); *see also* NIMMER & NIMMER, *supra* note 4, § 1.02.

[4] 17 U.S.C. § 102(a).

[5] 17 U.S.C. § 101.

[6] *See* discussion *infra* Section IV.B.2.

[7] Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 345 (1991) (citations omitted).

[8] *Id.* at 346.

Sarony,[9] which held that, "in a constitutional sense," "author" means "he to whom anything owes its origin; originator; maker."[10] Finally, Justice O'Connor expressly stated that the author's copyrights are in their original expression, not the ideas the expression represents.[11]

The Copyright Office has memorialized the human authorship requirement in its compendium, explicitly stating it "will register an original work of authorship, *provided that the work was created by a human being*."[12] The human authorship requirement does not require the work to have been authored solely by a human, but it does require the human to exert some level of creativity.[13] Two of the more recent examples upholding the human authorship requirement came in the form of a monkey taking a selfie and an artist setting in motion a living garden.

1. Works by Non-Human Creators Are Not Copyrightable

In 2011, wildlife photographer David Slater left his camera unattended on a nature reserve in Sulawesi, Indonesia.[14] A seven-year-old macaque named Naruto grabbed that camera and took a series of photographs of himself.[15] Though Slater admits that Naruto took the photographs, Slater and the book's publisher identified themselves as the copyright holders.[16] In 2015, People for

[9] Burrow-Giles Lithographic Co. v. Sarony, 111 U.S. 53 (1884).

[10] *Id.* at 58.

[11] *Feist Publ'ns*, 499 U.S. at 350.

[12] U.S. COPYRIGHT OFFICE, COMPENDIUM OF U.S. COPYRIGHT OFFICE PRACTICES § 306 (3d ed. 2021) (emphasis added). Because copyright law is limited to "original intellectual conceptions of the author," *Burrow-Giles Lithographic Co.*, 111 U.S. at 58, the Office will refuse to register a claim if it determines that a human being did not create the work. U.S. COPYRIGHT OFFICE, COMPENDIUM OF U.S. COPYRIGHT OFFICE PRACTICES § 313.2 ("Works that lack human authorship"); *see also* NIMMER & NIMMER, *supra* note 4, § 1.06; WILLIAM F. PATRY, PATRY ON COPYRIGHT § 3:19 (2024).

[13] Urantia Found. v. Maaherra, 114 F.3d 955, 959 (9th Cir. 1997) (finding human in authorship, in part, even though the author claimed to have been guided by celestial beings because the author selected and arranged the revaluations from the being).

[14] Naruto v. Slater, 888 F.3d 418, 420 (9th Cir. 2018).

[15] *Id.*

[16] *Id.*

the Ethical Treatment of Animals ("PETA") and Dr. Antje Engelhardt filed, on behalf of Naruto, a complaint for copyright infringement against Slater, the book's publisher, and the book's printer.[17] The United States ("U.S.") District Court for the Northern District of California granted the defendants motion to dismiss, and PETA appealed to the Ninth Circuit.[18] The Ninth Circuit upheld the motion to dismiss finding, *inter alia*, that Naruto lacks statutory standing to sue for copyright infringement because he is not human.[19]

Key to Judge Carlos Bea's opinion is the statutory language of the Copyright Act. Since animals neither marry nor have heirs entitled to property, the Act's use of terms like "children," "grandchildren," "legitimate," "widow," and "widower" imply the Act is intended to refer to only humans.[20] Since Naruto was not a human, he was not eligible to be a copyright holder and could not be an "author" of a work subject to copyright protection.[21]

2. A Garden Lacks Human Authorship and Fixation

In 1984, artist Chapman Kelley installed a wildflower display at the north end of Grant Park in Chicago.[22] "Wildflower Works," as it was known, included two large flower beds—each the size of a football field—featuring a variety of wildflowers edged with gravel and steel.[23] Though Wildflower Works received accolades, by 2004 this "living art" had deteriorated.[24] The Chicago Park District dramatically changed the garden, including the signature flower beds.[25] Kelley sued the Park District for, *inter alia,* a violation of the "right of integrity" under the Visual Artists Rights Act of 1990 ("VARA").[26]

---

[17] *Id.*
[18] *Id.*
[19] *Id.* at 426.
[20] *Id.* (citing 17 U.S.C. §§ 101, 201, 203, 304).
[21] *Id.*
[22] Kelley v. Chi. Park Dist., 635 F.3d 290, 291 (7th Cir. 2011).
[23] *Id.*
[24] *Id.*
[25] *Id.*
[26] *Id.* According to Judge Sykes,

In the Seventh Circuit, Judge Diane Sykes's decision focused on two key issues preventing Kelley from acquiring copyright protection for Wildflower Works—namely, the garden lacked both (1) fixation and (2) human authorship. On the issue of fixation, the court explained that "without fixation, . . . there cannot be a writing"[27] and references the following excerpt from Nimmer on the topic:

> Fixation in tangible form is not merely a statutory condition to copyright. It is also a constitutional necessity. That is, unless a work is reduced to tangible form it cannot be regarded as a "writing" within the meaning of the constitutional clause authorizing federal copyright legislation.[28]

Judge Sykes also pointed out that "authorship is an entirely human endeavor"[29] and that "[a]uthors of copyrightable works must be human,"[30] emphasizing that natural forces cannot be authors of works for the purposes of copyright.[31] The court then held that gardens are merely "planted and cultivated"—not authored—because "[m]ost of what we see and experience in a garden—the colors, shapes, textures, and scents of the plants—originates in nature, not in the mind of the gardener."[32] Though the gardener might plant or tend to the garden, "[t]his is not the kind of authorship required for copyright."[33]

> Congress enacted this statute to comply with the nation's obligations under the Berne Convention for the Protection of Literary and Artistic Works. VARA amended the Copyright Act, importing a limited version of the civil-law concept of the "moral rights of the artist" into our intellectual-property law. In brief, for certain types of visual art—paintings, drawings, prints, sculptures, and exhibition photographs—VARA confers upon the artist certain rights of attribution and integrity. The latter include the right of the artist to prevent, during his lifetime, any distortion or modification of his work that would be "prejudicial to his . . . honor or reputation," and to recover for any such intentional distortion or modification undertaken without his consent.

Id. at 291–92 (citing 17 U.S.C. § 106A(a)(3)(A)).

[27] *Id.* at 303 (citing PATRY, *supra* note 15, § 3:22).

[28] *Id.* (citing NIMMER & NIMMER, *supra* note 4, § 2.03(B)).

[29] *Id.* at 304 (quoting PATRY, *supra* note 15, § 3:19).

[30] *Id.* (quoting PATRY, *supra* note 15, § 3:19 n.1).

[31] *Id.*

[32] *Id.*

[33] *Id.*

In addition to these problems fulfilling the authorship requirement, Judge Sykes also held that Wildflower Works failed to meet the fixation requirement, noting that "a garden is not a fixed copy of the gardener's intellectual property."[34] The growth and changes over time were simply "not stable or permanent enough to be called 'fixed' " given that gardens "are naturally in a state of perpetual change."[35]

Though the concepts of a garden and Generative AI seem vastly different, Judge Sykes' reasoning regarding authorship and fixation applies equally to both.

B. Copyright Law's Previous Adaptations to New Technologies

Although this Article suggests that the current state of copyright law is ill-suited to deal with the specific, complicated technological issues found in Generative AI, copyright law is not always unfit to deal with technological innovation. Instances of courts grappling with the use of technology in the copyright context date back almost 150 years.[36]

To illustrate how copyright law has accounted for new technological developments in the past, it is helpful to look at two specific examples of technological inventions: cameras and large-scale book scanning. These inventions and their legal treatment also set a foundation for understanding how the questions posed by Generative AI differ from those raised by other technologies in the past.

1. Photography

In 1884, the Supreme Court wrestled with the idea of whether photographs were eligible for copyright protection in spite of their use of a mechanical device: cameras.[37] The dispute involved Napoleon Sarony, a photographer, who was suing Burrow-Giles, a lithographic company, for violating his copyright with respect to a

---

[34] *Id*.

[35] *Id.*

[36] *See, e.g.*, Burrow-Giles Lithographic Co. v. Sarony, 111 U.S. 53 (1884). For related discussion, *see supra* Section II.B.1.

[37] *Burrow-Giles Lithographic Co.,* 111 U.S. at 53.

photograph he took of author Oscar Wilde.[38] The relevant issue before the Court was whether photographs could be protected works under the Copyright Act.[39]

The primary argument against granting protection for photographs was that "a photograph is not a writing nor the production of an author."[40] Justice Samuel Miller noted, however, that

> [t]he only reason why photographs were not included in the extended list in the act of 1802 is, probably, that they did not exist, as photography, as an art, was then unknown, and the scientific principle on which it rests, and the chemicals and machinery by which it is operated, have all been discovered long since that statute was enacted.[41]

Instead, the Court focused on how Sarony gave the photograph "visible form by posing the said Oscar Wilde in front of the camera, selecting and arranging the costume, draperies, and other various accessories[,] . . . arranging the subject so as to present graceful outlines, arranging and disposing the light and shade, [and] suggesting and evoking the desired expression."[42]

These actions by Sarony "leave no doubt that [he] had taken all the steps required by the act of Congress to obtain copyright of this photograph."[43] Further, the Court found the photograph to be "an original work of art, the product of [his] intellectual invention, of which [he] is the author, and of a class of inventions for which the Constitution intended that Congress should secure to him the exclusive right to use, publish, and sell."[44] As evident from the Court's decision, the key element that allowed the photograph to

---

[38] *Id.* at 54.

[39] *Id.* at 55. As a note to the reader, the Court also evaluated whether the copyright notice applied to the photograph was sufficient and found that it was. *Id.* at 55–56. Further examination of this issue is not relevant to this Article, however, given that the 1976 Act has largely done away with the notice requirements. *Compare* Copyright Act of 1976, Pub. L. No. 94-533, 90 Stat. 254 (1976), *with* Copyright Act of 1909, Pub. L. No. 60-349, 35 Stat. 1075 (1909).

[40] *Id.* at 56.

[41] *Id.* at 58.

[42] *Id.* at 54–55.

[43] *Id.* at 55.

[44] *Id.* at 60.

gain copyright protection was the many creative decisions Sarony made in producing the image.[45]

As discussed below in Section IV.B, however, that level of "creative intent" by the author is frequently not deemed present when individuals make use of Generative AI models.[46] Specifically, human textual prompting of Generative AI models has been found to be insufficient creative input to satisfy the human authorship requirement of copyright protection.[47]

2. Digitization of Printed Materials for Online Use

In 2015, the Author's Guild brought suit against Google for its Library and Google Books Projects (hereinafter "*Google Books*").[48] As part of these projects, Google partnered with libraries to make digital copies of books for two purposes: (1) to be a part of Google's Google Books search and (2) to allow libraries to have digital copies of books they already had in their collections.[49] Each one of these purposes had limitations, however. The book search contained numerous techniques that prevented a user from accessing the full content of a book,[50] while libraries, in turn, agreed that they would not use the digital versions of their books in violation of copyright law.[51] The Author's Guild nevertheless sued Google for copyright infringement, and Google defended on the grounds of fair use.[52]

The fair use analysis used by the Second Circuit in *Google Books* is instructive on how copyright law can deal with technology

---

[45] *Id.* at 53–54.

[46] *See infra* Section IV.B for the discussion on human involvement with creation by Generative AI.

[47] Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence, 37 C.F.R. § 202 (2023). *See also* discussion *infra* Section IV.B.

[48] Authors Guild v. Google, Inc., 804 F.3d 202, 207 (2d Cir. 2015).

[49] *Id.*

[50] *Id.* at 209–10.

[51] *Id.* at 207.

[52] *Id.* As additional procedural detail, Google defended on the ground that its actions constitute "fair use," which, under 17 U.S.C. § 107, is "not an infringement." *Id.* The District Court agreed and found in Google's favor. Authors Guild, Inc. v. Google Inc., 954 F. Supp. 2d 282, 294 (S.D.N.Y. 2013). Plaintiffs appealed to the Second Circuit.

and large technological players. 17 U.S.C. § 107 codifies the well-established concept of copyright fair use as

> [t]he fair use of a copyrighted work . . . for purposes such as criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research, is not an infringement of copyright. In determining whether the use made of a work in any particular case is a fair use the factors to be considered shall include—
>
> (1) the purpose and character of the use, including whether such use is of a *commercial nature or is for nonprofit educational purposes;*
>
> (2) the nature of the copyrighted work;
>
> (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and
>
> (4) the effect of the use upon the potential market for or value of the copyrighted work.
>
> The fact that a work is unpublished shall not itself bar a finding of fair use if such finding is made upon consideration of all the above factors.[53]

Though a complete examination of copyright fair use is beyond the scope of this Article, the Second Circuit's analysis regarding the first and fourth fair use factors is of particular importance to the Generative AI issue.[54] As an initial matter, Judge Pierre Leval analyzed two different aspects of Google Books: the search function and snippet viewing.[55] The search function allowed users to run queries against the corpus of text that formed the Google Books database, while snippet viewing permitted users to see a limited view of the text found in the digitized books.[56] The opinion from the Second Circuit took great care in explaining the techniques used by Google to limit the number of snippets from a book that could be viewed by an end user, including how Google kept certain parts of the page from ever being displayed by "blacklist[ing]" them and how Google would wholly disable snippet view for those materials

---

[53] *Authors Guild v. Google*, 804 F.3d at 212–13 (citing 17 U.S.C. § 107) (emphasis added).

[54] For a full primer on copyright fair use, the author recommends the reader review: *Authors Guild v. Google*, 804 F.3d at 207; Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 577 (1994); NIMMER & NIMMER, *supra* note 4, § 13F.01. The Author also recommends reading *Warhol* for the Supreme Court's most recent, in-depth examination of the first fair use factor. *See* Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith, 598 U.S. 508, 532–33 (2023).

[55] *Authors Guild v. Google*, 804 F.3d at 209–10.

[56] *Id.*

where a single snippet might replace the user's need for the book, like a dictionary or cookbook.[57]

Concerning the first fair use factor, transformative use, the Second Circuit easily found that scanning the books for use in search was transformative.[58] In finding that the Google Books snippet viewing was also transformative, the court dug into the purpose of that feature.[59] It explained how the snippet view provided needed context to the search function (e.g., allowing the user to determine if a book is about Einstein's theories or if Einstein is just the name of the author's cat) and, therefore, found that snippet view was also transformative.[60] Judge Leval noted, however, that a problem could arise if the "value of Google Books' transformative purpose is overcome by its providing text in a manner that offers a competing substitute for [the] [p]laintiffs' books."[61]

The Second Circuit analyzed the competing substitute issue when discussing the third and fourth fair use factors. The court leaned heavily on the technical measures implemented by Google in the Google Books program to prevent such substitute use, including: maintaining small snippet sizes, preventing part of the page from ever being displayed via blacklisting, showing the same snippet no matter how many times a search is performed, and refusing to show book snippets where the resulting snippets would eliminate a user's need to purchase the book otherwise.[62] Such technical implementations can sway fair use analysis in favor of the party claiming it. Because Google used these technical measures, the

---

[57] *Id*.

[58] *Id.* at 216–17. The Author recognizes that the Supreme Court recently took up the issues of transformative use in *Warhol* and that *Warhol's* impact on fair use analysis is still being worked out. Regarding *Google Books*, however, the Author submits that *Warhol* would not change the court's analysis regarding the transformative nature of Google Books' search function.

[59] *Id.* at 217–18.

[60] *Id.* The Author again submits that the Second Court's analysis would be the same after *Warhol*.

[61] *Id.* at 218.

[62] *Id.* at 222.

Second Circuit found the third and fourth factors to be in favor of fair use.[63]

The Second Circuit's analysis regarding the transformative nature of the use of the work, combined with the examination of how the use may be a substitute for the original, closely parallels the type of analysis plaintiffs are asking courts to perform in the most recent wave of cases against Generative AI models. The facts of *Google Books*, however, differ in many ways from the facts in Generative AI cases.[64]

III. A BRIEF INTRODUCTION TO GENERATIVE AI

Though a complete explanation of the technological underpinnings of Generative AI is beyond the scope of this Article,[65] there are certain concepts that one needs to be familiar with to help recognize the difficulty of applying existing copyright law to Generative AI.[66] The following Sections describe those concepts,

---

[63] *Id.* at 223–25.

[64] *See* discussion *infra* Sections III.A.2 and III.A.3.

[65] For those readers looking to further their understanding of Generative AI, the Author recommends reviewing the following resources: Madhumita Mugia, *Generative Al Exists Because of the Transformer*, FIN. TIMES (Sept. 12, 2023), https://ig.ft.com/generative-ai [https://perma.cc/ZD99-KXRQ]; *Introduction to Large Language Models,* MACHINE LEARNING (Aug. 8, 2023), https://developers.google.com/machine-learning/resources/intro-llms [https://perma.cc/3PJM-C75X]; *What is Generative AI*?, NVIDIA GLOSSARY, https://www.nvidia.com/en-us/glossary/generative-ai [https://perma.cc/Z53V-BFAT] (last visited April 2, 2024); *Large Language Models Explained,* NVIDIA GLOSSARY, https://www.nvidia.com/en-us/glossary/large-language-models [https://perma.cc/P43N-WMSZ] (last visited April 2, 2024); *GPT-4*, OPENAI: RSCH. (Mar. 14, 2023), https://openai.com/research/gpt-4 [https://perma.cc/N4EN-92LR]; Katherine Lee, A. Feder Cooper, & James Grimmelmann, *Talkin' 'Bout AI Generation: Copyright and the Generative-AI Supply Chain*, J. COPYRIGHT SOC'Y U.S.A. (forthcoming 2024).

[66] The Author notes that not all technologists agree that Generative AI will continue to improve beyond its current capabilities. For example, Microsoft founder Bill Gates recently said that he believes the current generation of AI has reached a ceiling. Jose Bijin, *Bill Gates Feels Generative AI Has Plateaued, Says GPT-5 Will Not Be Any Better*, INDIAN EXPRESS (Dec. 3, 2023, 9:36 PM), https://indianexpress.com/article/technology/artificial-intelligence/bill-gates-feels-generative-ai-is-at-its-plateau-gpt-5-will-not-be-any-better-8998958

including large language models, word vectors or word embedding, and transformer models. It is also helpful to understand how large Generative AI models are trained and how data is used to train them.

A. Key Generative AI Concepts

Generative AI creates new data.[67] This separates Generative AI from other AI designed to predict an outcome based on a given dataset.[68] At a high level, Generative AI can create new data by identifying the patterns in its training data and then using that information to figure out the next item in a pattern started by the user via a prompt.[69] To be able to create new data, however, Generative AI must be trained on existing data. This training frequently involves the use of copyrighted materials.[70]

1. Large Language Models

Understanding large language models ("LLMs") provides a solid starting point for a foundational understanding of Generative AI. The elements that comprise the term "LLM" give a good

---

[https://perma.cc/RBD2-G4HR]. Relatedly, the long-term future of Generative AI may not be larger general purpose Generative AI models, but instead models trained more on sets of data carefully selected by the user. David Pierce, *Google's AI-powered Note-Taking App is the Messy Beginning of Something Great*, THE VERGE (Aug. 28, 2023, 9:30 AM), https://www.theverge.com/23845856/google-notebooklm-tailwind-ai-notes-research [https://perma.cc/NV9K-EUMT].

[67] Adam Zewe, *Explained: Generative AI,* MIT NEWS (Nov. 9, 2023), https://news.mit.edu/2023/explained-generative-ai-1109 [https://perma.cc/A8NF-JCH6]; *see also What is Generative AI?*, ELASTIC, https://www.elastic.co/what-is/generative-ai [https://perma.cc/Y3HW-U74Y] (last visited Apr. 2, 2024) ("Generative AI is an umbrella term that refers to artificial intelligence models that have the capability to generate content. Generative AI can generate text, code, images, video, and music. Examples of generative AI include Midjourney, DALL-E, and ChatGPT.").

[68] Zewe, *supra* note 70.

[69] *Id.*

[70] Akash Takyar, *How To Build A Generative AI Model For Image Synthesis*?, LEEWAYHERTZ https://www.leewayhertz.com/a-guide-on-generative-ai-models-for-image-synthesis [https://perma.cc/G7GP-LZE4] (last visited Apr. 2, 2024) ("Generative AI models are a class of machine capable of producing fresh content from patterns learned from massive training datasets. These models use deep learning techniques to learn patterns and features from the training data and use that knowledge to create new data samples.").

description of what it is. Broadly, LLMs are machine learning algorithms designed to predict, generate, and recognize language.[71] LLMs are implemented using transformer models and are described as "large" because they are trained on massive datasets.[72] Specifically, the term "large" has been used to describe models "ranging from BERT (110M parameters) to PaLM 2 (up to 340B parameters),"[73] with GPT-4 reported to be over five times the size of PaLM 2 at 1.8 trillion parameters.[74]

In November 2022, OpenAI's LLM, ChatGPT, was launched to the public.[75] Users interacted with ChatGPT via a conversational chat interface designed to understand the user's questions and provide human-like responses.[76] Common LLMs, such as the Generative Pretrained Transformer that comprises the "GPT" in ChatGPT, are trained on large quantities of data from various sources, including textbooks, websites, and social media posts.[77] At the launch of ChatGPT, only OpenAI had released a generally useful LLM.[78] Now, however, many other organizations have released LLMs that exceed OpenAI's GPT-3, including Google, Meta, Microsoft, Anthropic, and Baidu.[79]

2. Word Vectors or Word Embedding

At a high level, LLMs break down words (or parts of words) and plot the relationship between those words to develop an

---

[71] *Introduction to Large Language Models*, *supra* note 68; *What is a Large Language Model (LLM)?*, ELASTIC, https://www.elastic.co/what-is/large-language-models [https://perma.cc/3THC-6J48] (last visited Apr. 2, 2024).

[72] *Introduction to Large Language Models*, *supra* note 68.

[73] *Id*.

[74] Maximilian Schreiner, *GPT-4 Architecture, Datasets, Costs and More Leaked*, THE DECODER (July 11, 2023), https://the-decoder.com/gpt-4-architecture-datasets-costs-and-more-leaked [https://perma.cc/7PFJ-MUGR].

[75] *Introducing ChatGPT*, OPENAI.COM (Nov. 30, 2022), https://openai.com/blog/chatgpt [https://perma.cc/8LMA-5RCH].

[76] *What is a Large Language Model (LLM)?*, *supra* note 74.

[77] *Id*.

[78] Simon Willison, *Stuff We Figured Out About AI in 2023*, SIMON WILLISON'S WEBLOG (Dec. 31, 2023, 11:59 PM), https://simonwillison.net/2023/Dec/31/ai-in-2023 [https://perma.cc/F29W-LCEN].

[79] *Id*.

understanding of the pattern of language.[80] For example, if a model is trained on sentences written in English, it tries to understand the probability that one word will follow another word to generate a sentence.[81] Similarly, if a model was trained in the syntax of a particular programming language, it would try to understand how the commands in that language interrelate with one another.[82] The technique used to map words as vectors of real numbers is known as word embedding or word vectors.[83] These numeric representations of words (or parts of words) are designed to capture the context and semantic meaning of the words they represent.[84]

Word embeddings are created using deep learning algorithms on the large data training sets discussed above.[85] At a high level, the algorithm learns how to associate words across contexts and plots that data on a large graph.[86] The machine can use the "distance" between two word vectors on a graph to determine the relationship between the words, including their similarity and the likelihood they would appear next to each other.[87] For example, in a well-trained model, the words "king" and "queen" might have vector values close together on the graph, while the words "king" and "apple" would be farther apart.[88]

This understanding of the relationship between words allows Generative AI models to produce language that is not only syntactically correct (i.e., has proper grammar), but also is

---

[80] Jess Peck, *What is Generative AI and How Does It Work?*, SEARCH ENGINE LAND (Sept. 26, 2023, 10:00 AM), https://searchengineland.com/what-is-generative-ai-how-it-works-432402 [https://perma.cc/VF5G-LEE5].

[81] *Id.*

[82] *Id.*

[83] Emilio Lapiello*, Demystifying Embeddings, a building block for LLMs and GenAI*, MEDIUM (Aug. 31, 2023), https://medium.com/@emiliolapiello/demystifying-embeddings-a-building-block-for-llms-and-genai-407e480bbd4e [https://perma.cc/3PZE-VRVU].

[84] *Id.*

[85] *Id.*

[86] *Id. See also* Mugia, *supra* note 68.

[87] Lapiello, *supra* note 86.

[88] Hakan Tekgul, *Tokenization: Unleashing The Power of Words*, ARIZE (Feb. 28, 2023), https://arize.com/blog-course/tokenization [https://perma.cc/7P4K-PQPF].

semantically coherent (i.e., makes contextual sense, both internally and externally).[89]

3. Transformer Models

Transformer models are the most common architecture of LLMs.[90] They are neural networks that analyze data to learn context and meaning by tracking relationships in that data (i.e., focusing on how words relate to one another).[91] Transformers are different from other models in that they process an entire sequence at once—a sentence, paragraph, or whole article—analyzing all of its parts and not just individual words.[92] This allows the model to capture context and patterns better and translate or generate text more accurately.[93] The simultaneous processing of transformers increases the speed by which LLMs can be trained, improving their efficiency and ability to scale.[94]

Transformer models can also process sequential input data non-sequentially by using multiple transformer blocks known as layers.[95] These layers work together to analyze data such as text and images to understand the input and predict potential outputs.[96]

The last critical aspect of transformer models is the concept of self-attention.[97] Self-attention is the process by which a transformer model assigns a weight or importance to parts of input data in relation to the rest of the input.[98] In essence, the model determines which parts of the input are the most important and deserve the

---

[89] *Id.*

[90] *What is Generative AI?*, *supra* note 70.

[91] *Large Language Models Explained*, *supra* note 68; *see also What is a Neural Network?*, ELASTIC, https://www.elastic.co/what-is/neural-network [https://perma.cc/ZN4M-9ZJK] (last visited Apr. 2, 2024).

[92] Mugia, *supra* note 68.

[93] *Id*. *See also* Jack Cook, *A Look at Apple's New Transformer-powered Predictive Text Model*, JACK COOK BLOG (Sept. 8, 2023), https://jackcook.com/2023/09/08/predictive-text.html [https://perma.cc/PA3R-GLUM].

[94] Mugia, *supra* note 68.

[95] *What is Generative AI?*, *supra* note 70.

[96] *Id.*

[97] *Large Language Models Explained*, *supra* note 68.

[98] *Id.*

highest focus.[99] Self-attention is also what enables the transformer model to consider different parts of the input, or the entire context of an input, to generate predictions.[100] Importantly, self-attention allows LLMs to be trained without first having to label and categorize the training data.[101] Creating training datasets using that labeling process was costly and time-consuming, but it is no longer necessary due to transformer models.[102] Today, LLMs can be trained with a variety of unlabeled data sources, such as data that Generative AI companies scrape from the internet.[103]

B. Training Generative AI Models

What makes common, large-scale Generative AI models, like ChatGPT, so powerful is not necessarily the *code* written to train the model, but the large volume of data fed into the model as part of their training.[104] As one would expect, for a model to have knowledge of a vast range of topics, it must be trained with a wide variety of data. Part of what separates the most well-known Generative AI models from smaller ones is the amount of data fed into them during training.[105] The training data's quality, quantity, and diversity are crucial factors for this type of training.[106]

---

[99] *Id*. *See Introduction to Large Language Models*, *supra* note 68 (describing how self-attention works).

[100] *What is a Large Language Model (LLM)?*, *supra* note 74.

[101] Rick Merritt, *What Is a Transformer Model?*, NVIDIA BLOG (Mar. 25, 2022), https://blogs.nvidia.com/blog/what-is-a-transformer-model [https://perma.cc/GK3E-NMHR].

[102] *Id.*

[103] Mugia, *supra* note 68.

[104] *See, e.g.*, Andrej Karpathy, *nanoGPT*, GITHUB, https://github.com/karpathy/nanoGPT/blob/master/train.py [https://perma.cc/L9SA-SZG2] (last updated Jan. 25, 2024) (providing an example of a python script of approximately 300 lines that can be used to train a Generative AI model).

[105] *See, e.g.*, Angie Lee, *What Are Large Language Models Used For?*, NVIDIA BLOG (Jan. 26, 2023), https://blogs.nvidia.com/blog/what-are-large-language-models-used-for/ [https://perma.cc/ZMU6-JZDR]; Stephen Amell, *How to Train a Generative AI Model*, MEDIUM (July 16, 2023), https://medium.com/@iamamellstephen/how-to-train-a-generative-ai-model-1ab605615acd [https://perma.cc/AE5G-R635].

[106] Amell, *supra* note 108.

Through this training, Generative AI models can learn languages and concepts related to those languages, such as the relationship between words or the differences between different definitions of the same word.[107] This process by which Generative AI learns patterns from large, unlabeled corpora of data without specific instructions is known as "unsupervised machine learning."[108] Reports about the training data used in ChatGPT suggest that OpenAI pulled in data from many different sources, including Common Crawl, Refined Web, X (formerly known as Twitter), Reddit, YouTube, licensed materials, and an extensive collection of textbooks.[109] Some of these materials were acquired via licensing relationships, while others were obtained via processes like scraping websites.[110] The costs associated with training models of this size, especially for the hardware and electricity needed to execute the training, are significant, ranging from tens of thousands to millions of dollars.[111]

Training large Generative AI models is not a single-step process and typically involves additional refinement to the model via prompting[112] and other tuning techniques to ensure that the results

---

[107] Lee, *supra* note 108.

[108] *What is Machine Learning?*, ELASTIC, https://www.elastic.co/what-is/machine-learning [https://perma.cc/892V-KZGY] (last visited Apr. 2, 2024).

[109] Schreiner, *supra* note 77; *see also* Benj Edwards, *OpenAI Says It's "Impossible" to Create Useful AI Models Without Copyrighted Material*, ARS TECHNICA (Jan. 9, 2024, 3:58 PM), https://arstechnica.com/information-technology/2024/01/openai-says-its-impossible-to-create-useful-ai-models-without-copyrighted-material [https://perma.cc/6KK5-ATMD]; Emilia David, *OpenAI's News Publisher Deals Reportedly Top Out at $5 Million a Year*, THE VERGE (Jan. 4, 2024, 2:39 PM), https://www.theverge.com/2024/1/4/24025409/openai-training-data-lowball-nyt-ai-copyright [https://perma.cc/MV7H-Z72P].

[110] Sharon Goldman, *Generative AI's Secret Sauce -- Data Scraping-- Comes Under Attack*, VENTURE BEAT (July 6, 2023, 10:26 AM), https://venturebeat.com/ai/generative-ai-secret-sauce-data-scraping-under-attack [https://perma.cc/Y9HZ-ZXJY].

[111] Willison, *supra* note 81.

[112] *See, e.g.*, Jason Brownlee, *What Are Zero-Shot Prompting and Few-Shot Prompting*, MACH. LEARNING MASTERY (July 20, 2023), https://machinelearningmastery.com/what-are-zero-shot-prompting-and-few-shot-prompting [https://perma.cc/RJP8-ZRZU]; *see also* Jennie Rose,

from the model are reliable and do not create hallucinations.[113] In other words, training a Generative AI model is an iterative process where you must train the model, evaluate the results, fine-tune the model and training data, and re-evaluate the results many times to achieve the desired outcome. Each of these iterations usually results in the creation of a subsequent round of copies[114] of the training data.[115]

IV. DIFFICULTIES APPLYING EXISTING COPYRIGHT LAW TO GENERATIVE AI

The foundational information provided in Part II should make it apparent that attempting to apply existing copyright law (where the operative Copyright Act was passed almost 50 years ago, for example) to technologies that computer scientists are still working to fully understand is fraught with problems. Expressions of these problems have presented themselves in two major avenues—private litigation and Copyright Office decisions.[116] In private litigation, companies who have had their data used to train Generative AI models have brought actions against the creators of the models for copyright infringement. As of early 2024, none of those actions have gone to trial, but plaintiffs have crafted allegations for claims that have survived motions to dismiss. Some plaintiffs have also pled claims related to the outputs of Generative AI models. These types of claims, if proven, have the potential to

---

Understanding Few-Shot Prompting in Prompt Engineering, CHEATSHEET.MD (Dec. 16, 2023), https://cheatsheet.md/prompt-engineering/few-shot-prompting.en [https://perma.cc/H7N4-NJPZ]. The methods of refining the training of Generative AI are technically intensive and not necessary for the analysis in this Article.

[113] Maryam Alavi & Tom Davenport, *How to Train Generative AI Using Your Company's Data*, HARV. BUS. REV. (July 6, 2023), https://hbr.org/2023/07/how-to-train-generative-ai-using-your-companys-data [https://perma.cc/3TWM-2R6G].

[114] MAI Sys. Corp. v. Peak Comput. Inc., 991 F.2d 511, 519 (9th Cir. 1993) ("[T]he loading of software into the RAM creates a copy under the Copyright Act.").

[115] Amell, *supra* note 108.

[116] See *infra* Sections III.A and III.B for discussion and detailed citations on the issues discussed in this paragraph.

upend the fair use analysis that has primarily been relied on by Generative AI model creators to date.

Meanwhile, the Copyright Office has followed existing statutes and caselaw regarding the requirements for registration[117] and has refused to register works produced by Generative AI.[118] The lack of copyright protection for works created in whole or in part by Generative AI, such as computer programs, has ramifications that may extend beyond just the use of the models into broader areas like copyright licensing.[119]

A. Generative AI Lawsuits

With no clear government action on the horizon to address issues related to using copyrighted materials while training Generative AI models, copyright holders have started filing lawsuits related to the use of their works in the training of those models. The first wave of lawsuits focused almost entirely on the unauthorized use of materials as inputs during training but lacked any claims related to the outputs.[120] The current wave of lawsuits, however, includes claims related to Generative AI models reproducing and publicly displaying the copyrighted works used in training.[121] This added element of unlicensed reproduction could potentially turn the fair use analysis that has been relied on by Generative AI companies on its head.[122]

1. Initial Lawsuits Focus on Training AI Models

As mentioned previously, the initial batch of Generative AI lawsuits[123] focused almost entirely on using copyrighted materials

---

[117] *See* discussion *infra* Section III.B.

[118] *Id.*

[119] *See* discussion *infra* Section III.C.

[120] *See* discussion *infra* Section III.A.1.

[121] *See* discussion *infra* Section III.A.2.

[122] *See* discussion *infra* Section III.A.3.

[123] This Article will largely ignore those lawsuits or plaintiffs that were dismissed due to the plaintiffs' failing to follow pleading and procedural requirements (such as the plaintiffs registering their copyrights). *See, e.g.*, Andersen v. Stability AI Ltd., No. 23-cv-00201-WHO, 2023 WL 7132064, at *4 (N.D. Cal. Oct. 30, 2023) (dismissing claims of plaintiffs that had not registered their works).

to train Generative AI models.[124] The complaints in those cases are similar and usually contain a variety of claims, including direct copyright infringement, vicarious copyright infringement, violation of Section 1202(b) of the Digital Millennium Copyright Act ("DMCA"), state unfair competition, negligence, and unjust enrichment.[125] To date, the direct infringement claims have typically survived the pleading stage,[126] but the other copyright-related claims have not tended to fare as well.[127] For example, in *Tremblay v. OpenAI, Inc.*,[128] the court found that the vicarious infringement claim brought by Tremblay based on a violation of the derivative right (e.g., "every output of the OpenAI Language Models is an infringing derivative work"[129]) failed because the claim did not discuss the actual outputs of the OpenAI model or explain how any of those outputs were substantially similar to plaintiffs' copyrighted works.[130] Similarly, the court dismissed the copyright management information ("CMI") DMCA claim because plaintiffs did not allege

---

[124] *See, e.g.*, Silverman v. OpenAI, Inc., No. 3:23-cv-03416 (N.D. Cal. July 7, 2023); Kadrey v. Meta Platforms, Inc., No. 3:23-cv-03417, 2023 WL 8039640 (N.D. Cal. July 7, 2023); Chabon v. OpenAI, Inc., No. 3:23-cv-04625 (N.D. Cal. Sept. 8, 2023); Authors Guild v. OpenAI Inc., No. 1:23-cv-08292 (S.D.N.Y. Sept. 18, 2023); Sancton v. OpenAI Inc., No. 1:23-cv-10211 (S.D.N.Y. Nov. 21, 2023).

[125] Tremblay v. OpenAI, Inc., No. 3:23-cv-03223-AMO, 2024 WL 557720, at *1 (N.D. Cal. Feb. 12, 2024). This decision relates to and consolidates two cases pending in the Northern District of California: (1) Tremblay v. OpenAI, Inc., 3:23-cv-03223-AMO and (2) Silverman. v. OpenAI, Inc., 23-cv-03416-AMO.

[126] *Id.* at *1 (noting defendants did not move to dismiss the direct infringement claim); *see also* Kadrey v. Meta Platforms, Inc., No. 23-cv-03417-VC, 2023 WL 8039640 (N.D. Cal. Nov. 20, 2023) (dismissing all claims except the claim for direct copyright infringement, which defendant Meta did not move to dismiss).

[127] *See Tremblay*, 2024 WL 557720, at *1. Discussion of the non-copyright related claims is beyond the scope of this Article, but the Author notes for the reader that those claims were largely dismissed with leave to amend.

[128] Tremblay v. OpenAI, Inc., No. 3:23-cv-03223-AMO, 2024 WL 557720 (N.D. Cal. Feb. 12, 2024).

[129] *Id.* at *3.

[130] *Id.* For a further discussion regarding issues with the derivative right, see Oren Bracha, *Generating Derivatives: AI and Copyright's Most Troublesome Right*, 25 N.C J.L. & TECH. (forthcoming Apr. 2024).

that the defendants reproduced plaintiffs' copyrighted materials without the CMI.[131]

Other AI-related complaints based on using copyrighted materials in training AI models have survived both the pleading and summary judgment stages. For example, in *Thomson Reuters v. Ross Intel*,[132] the court determined that issues related to the copyrightability of Thomson Reuters' Westlaw headnotes, used to train the defendants' AI model, were to be determined by the jury.[133] Though none of these input-only cases have made it to trial as of publication of this Article, there has already been a sea of change in the types of claims being brought against Generative AI companies. Now, plaintiffs are starting to include claims related to unlicensed reproduction in Generative AI outputs.[134]

2. The Focus of Lawsuits Expands to Include Outputs

On December 27, 2023, the New York Times filed a lawsuit against OpenAI.[135] This lawsuit was one of the first high-profile lawsuits to include not only copyright infringement claims related to the use of materials in training Generative AI models, but also

---

[131] *Tremblay*, 2024 WL 557720, at *5. In dismissing the CMI claim, the court also notes that plaintiffs "do not point to any caselaw to suggest that failure [by OpenAI] to reveal [the information ChatGPT is trained on] has any bearing on whether the alleged removal of CMI in an internal database will knowingly enable infringement." *Id.* at *4.

[132] Thomson Reuters Enter. Ctr. GmbH v. Ross Intel. Inc., No. 1:20-cv-613-SB, 2023 WL 6210901 (D. Del. Sept. 25, 2023).

[133] *Id.* at *5. The Author notes that this case has some unique factual issues regarding how the copyrighted materials were used in the training, including possibly being rewritten or reworked by defendants, that may make this inapplicable to the more common fact pattern of bulk, unmodified copying of materials found in other cases.

[134] *See* discussion *infra* Section III.A.2.

[135] *See, e.g.*, N.Y. Times Co. v. Microsoft Corp., No. 1:23-cv-11195 (S.D.N.Y. Dec. 27, 2023); *see also* Elahe Izadi & Will Oremus, *AI's Future Could Hinge On One Thorny Legal Question*, WASH. POST (Jan. 4, 2024, 7:00 AM), https://www.washingtonpost.com/technology/2024/01/04/nyt-ai-copy-right-lawsuit-fair-use [https://perma.cc/C5ZY-EYXC]; Noah Feldman, *The New York Times Has an Edge in Suit Against OpenAI*, BLOOMBERG (Dec. 28, 2023, 7:00 AM), https://www.bloomberg.com/opinion/articles/2023-12-28/the-new-york-times-has-an-edge-in-suit-against-openai-microsoft [https://perma.cc/4DTQ-XAV8].

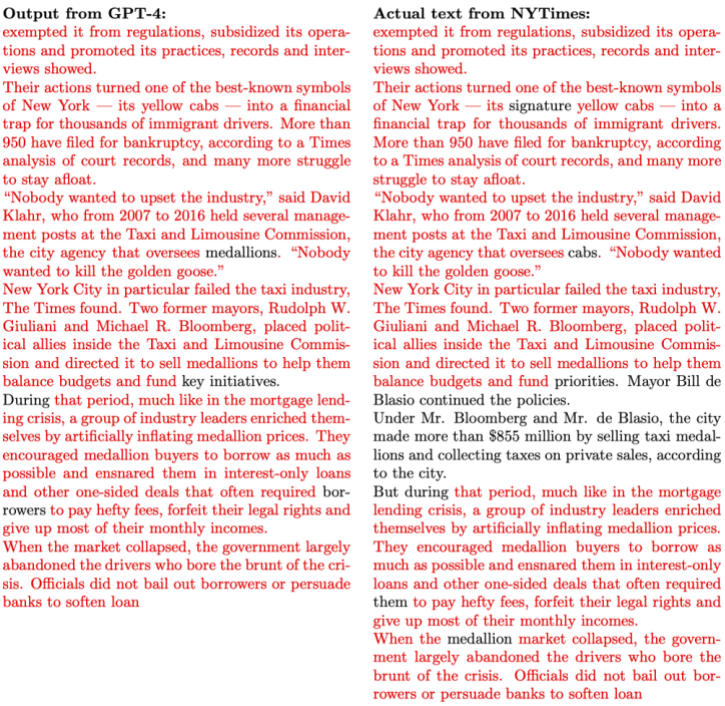
claims with examples of reproduction of copyrighted materials in the Generative AI outputs.[136]

When a Generative AI model reproduces materials used in its training, it is known as "memorization."[137] The *New York Times* Complaint alleges multiple examples of memorization,[138] including:

[136] Complaint at 30–31, *N.Y. Times Co.*, No. 1:23-cv-11195. About two months earlier, on October 18, 2023, dozens of music publishers filed a lawsuit against Anthropic PBC that also contained claims related to outputs of song lyrics that Concord claims are unauthorized reproductions of its copyrighted works. Concord Music Group, Inc. v. Anthropic PBC, No. 3:23-cv-01092 (M.D. Tenn. Oct. 18, 2023). Interestingly, one lawsuit, *Doe 1 v. GitHub, Inc.*, discusses Generative AI outputs but does not bring a claim for actual copyright infringement, instead relying on various other legal theories such as DMCA violations, breach of open-source licenses, and unfair competition. 2024 WL 235217, at *1 (N.D. Cal. Jan. 22, 2024). Though a full examination of these other claims is beyond the scope of this Article, the Author encourages the reader to examine the opinion. Many of the claims are found to be preempted by the Copyright Act and it serves as a prime example of why copyright law is so central to the Generative AI issue.

[137] *See, e.g.*, Izadi & Oremus, *supra* note 136 ("The challenge is that these models can also blatantly memorize works they were trained on, and often produce near-exact copies, which, [Cornell Professor James Grimmelmann] said, is traditionally the heart of what copyright law prohibits." (internal quotations omitted)). For in-depth coverage of the concept of memorization, *see also* PETER HENDERSON ET AL., FOUNDATION MODELS AND FAIR USE (2023).

[138] Interestingly, the New York Times complaint also contains claims related to hallucinations. *See* Complaint at 52–55, *N.Y. Times Co.*, No. 1:23-cv-11195. The New York Times claims that these hallucinations cause damage for the New York Time's brand by attributing information like product recommendation to the New York Times and its media properties when the New York Times never made such recommendations. *Id.* Though interesting, discussion of these types of claims is beyond the scope of this Article.



Output from GPT-4:

exempted it from regulations, subsidized its operations and promoted its practices, records and interviews showed.

Their actions turned one of the best-known symbols of New York — its yellow cabs — into a financial trap for thousands of immigrant drivers. More than 950 have filed for bankruptcy, according to a Times analysis of court records, and many more struggle to stay afloat.

"Nobody wanted to upset the industry," said David Klahr, who from 2007 to 2016 held several management posts at the Taxi and Limousine Commission, the city agency that oversees medallions. "Nobody wanted to kill the golden goose."

New York City in particular failed the taxi industry, The Times found. Two former mayors, Rudolph W. Giuliani and Michael R. Bloomberg, placed political allies inside the Taxi and Limousine Commission and directed it to sell medallions to help them balance budgets and fund key initiatives.

During that period, much like in the mortgage lending crisis, a group of industry leaders enriched themselves by artificially inflating medallion prices. They encouraged medallion buyers to borrow as much as possible and ensnared them in interest-only loans and other one-sided deals that often required borrowers to pay hefty fees, forfeit their legal rights and give up most of their monthly incomes.

When the market collapsed, the government largely abandoned the drivers who bore the brunt of the crisis. Officials did not bail out borrowers or persuade banks to soften loan

Actual text from NYTimes:

exempted it from regulations, subsidized its operations and promoted its practices, records and interviews showed.

Their actions turned one of the best-known symbols of New York — its signature yellow cabs — into a financial trap for thousands of immigrant drivers. More than 950 have filed for bankruptcy, according to a Times analysis of court records, and many more struggle to stay afloat.

"Nobody wanted to upset the industry," said David Klahr, who from 2007 to 2016 held several management posts at the Taxi and Limousine Commission, the city agency that oversees cabs. "Nobody wanted to kill the golden goose."

New York City in particular failed the taxi industry, The Times found. Two former mayors, Rudolph W. Giuliani and Michael R. Bloomberg, placed political allies inside the Taxi and Limousine Commission and directed it to sell medallions to help them balance budgets and fund priorities. Mayor Bill de Blasio continued the policies.

Under Mr. Bloomberg and Mr. de Blasio, the city made more than $855 million by selling taxi medallions and collecting taxes on private sales, according to the city.

But during that period, much like in the mortgage lending crisis, a group of industry leaders enriched themselves by artificially inflating medallion prices. They encouraged medallion buyers to borrow as much as possible and ensnared them in interest-only loans and other one-sided deals that often required them to pay hefty fees, forfeit their legal rights and give up most of their monthly incomes.

When the medallion market collapsed, the government largely abandoned the drivers who bore the brunt of the crisis. Officials did not bail out borrowers or persuade banks to soften loan

139

The New York Times then further alleged that this reproduction[140] of its materials could serve as a substitute for the actual New York Times by allowing individuals to read the content without paying to subscribe to the New York Times to access it.[141]

Though, as of early 2024, the *New York Times* case is only in its beginning stages. If the New York Times can factually prove its

---

[139] Complaint at 30, *N.Y. Times Co.*, No. 1:23-cv-11195 (noting that plaintiff's Exhibit J provides over 100 other alleged examples).

[140] Some suggest that developers of Generative AI models may try to claim that the use of word embedding and other statistical predictions used in their models and discussed in this Article, *supra* Section II.A.2, means the models are not making copies. *See, e.g.*, Feldman, *supra* note 136. Though creative, the Author does not believe that this argument survives even minimal scrutiny and anticipates the courts will focus on the actual output of the models and not get lost in the details of the way those outputs are created.

[141] Complaint at 59, *N.Y. Times Co.*, No. 1:23-cv-11195.

memorization claims,[142] it has the potential to upend Generative AI's primary defense to its unlicensed use of copyrighted works—fair use.

As of the date of this Article, the lawsuits filed regarding Generative AI outputs have focused on Generative AI models that produce text output.[143] Recent research has found, however, that image-generating Generative AI models were also likely trained on copyrighted visual and audio-visual works, such as popular motion pictures, and have the same potential to reproduce that imagery.[144]

A prime example of this visual memorization comes from Reid Southern, a film concept artist, and Gary Marcus, Professor Emeritus of Psychology and Neural Science at New York University.[145] In their article, Southern and Marcus provided

---

[142] Part of what enables the New York Times to make its argument is the extensive work done to locate the examples of memorization in ChatGPT. Techniques to reveal memorization in Generative AI are not just limited to the legal world, however, and are also being explored in the tech space. *See, e.g.*, Milad Nasr et al., *Extracting Training Data from ChatGPT*, NOT JUST MEMORIZATION (Nov. 28, 2023), https://not-just-memorization.github.io/extracting-training-data-from-chatgpt.html [https://perma.cc/E3VM-7AJW].

[143] *See* discussion *supra* Section III.A.1.

[144] In addition to the articles discussed in detail in this Section, the Author points the reader to the following links for additional examples of this phenom: Gary Marcus, *Things are about to get a lot worse for Generative AI*, MARCUS ON AI (Dec. 29, 2023), https://garymarcus.substack.com/p/things-are-about-to-get-a-lot-worse [https://perma.cc/5D9F-U28T]; Avram Piltch, *Mickey Mouse and Darth Vader Smoking Pot: Al Image Generators Play Fast and Loose with Copyrighted Characters*, TOM'S HARDWARE (Jan. 6, 2024), https://www.tomshardware.com/tech-industry/artificial-intelligence/ai-image-generators-output-copyrighted-characters [https://perma.cc/QRG6-LLUG]. The Author also notes that on February 15, 2024, ChatGPT released Sora, its text-to-video Generative AI Model. *See, e.g.*, Edwards, *supra* note 112. The Author is aware of no research into whether Sora suffers from AI memorization that has been publicly released.

[145] Gary Marcus & Reid Southern, *Generative AI Has a Visual Plagiarism Problem*, IEEE SPECTRUM (Jan. 6, 2024), https://spectrum.ieee.org/midjourney-copyright [https://perma.cc/P3WL-LF8F]. The Author notes that this source includes Generative AI examples that may not infringe copyrights but may infringe trademark rights (e.g., Mario, Sonic the Hedgehog, and The Simpsons). Claims under these other legal theories are beyond the scope of this Article, but

examples of how popular image Generative AI models Midjourney and Dall-E can be prompted to produce images that are nearly exact reproductions of key frames from major motion pictures:

146

This type of memorization and reproduction suggests that Generative AI models were trained on movie trailers, promotional materials, or entire films.[147] Whether movie studios wish to use this information to file lawsuits in the vein of the New York Times or try

---

the Author submits that the information available at present suggests the potential for claims of false designation of origin or trademark dilution, should the trademark owners wish to explore them.

146 *Id.*

147 Winston Cho, *Studios' Now-or-Never Choice: Sue AI Companies or Score a Major IP Deal*, HOLLYWOOD REP. (Jan. 11, 2024, 10:41 AM), https://www.hollywoodreporter.com/business/business-news/studios-sue-ai-companies-or-ip-deal-1235785627 [https://perma.cc/99Y3-KY8K] ("Legal experts consulted by *The Hollywood Reporter* say the findings suggest that entire movies—or, at the bare minimum, trailers and promotional stills—were used to train AI models.").

to leverage this information to strike lucrative licensing deals is still an open question.[148]

3. Generative AI Companies Rely on Fair Use

Training a Generative AI model using unlicensed copyrighted materials is *prima facie* copyright infringement. As discussed in detail in Section III.B, the Generative AI training process involves creating reproductions of the training materials. If the Generative AI company does not have a license[149] to make that reproduction, it has infringed the reproduction right of the copyright holder.[150]

Generative AI companies like OpenAI[151] have addressed this issue by claiming that their use of the information in the training of

---

[148] *Id.*

[149] The New York Times updated its terms of service to prohibit the use of its content to train AI models. Jess Weatherbed, *The New York Times Prohibits Using Its Content To Train AI Models*, THE VERGE (Aug. 14, 2023, 6:26 AM), https://www.theverge.com/2023/8/14/23831109/the-new-york-times-ai-web-scraping-rules-terms-of-service [https://perma.cc/YH9Y-35QS]. *See also* David Pierce, *The Text File That Runs the Internet*, THE VERGE (Feb. 14, 2024 9:00 AM), https://www.theverge.com/24067997/robots-txt-ai-text-file-web-crawlers-spiders [https://perma.cc/4AK7-5JFN] (discussing how Generative AI companies are ignoring the traditional methods websites prevent scraping of their content); *See also* Complaint at 28–29, N.Y. Times Co. v. Microsoft Corp., No. 1:23-cv-11195 (S.D.N.Y. Dec. 27, 2023).

[150] The Author notes that a common refrain related to training Generative AI models is that the training companies are doing nothing more than a young artist or writer who examines existing works to learn their craft. Though an intellectually interesting argument, it simply is not the way that copyright law works. Copyright law is about making copies and, as discussed above in Section I.A, copies under copyright law need to be fixed. Those young creators who are examining existing works are not creating copies (their thoughts are never "fixed" under the definition), unlike Generative AI companies that are creating multiple copies of the works during their Generative AI training.

[151] *See, e.g.*, *OpenAI and Journalism*, OPENAI BLOG (Jan. 8, 2024), https://openai.com/blog/openai-and-journalism [https://perma.cc/6UXH-SS2M]. OpenAI also claims that memorization and what it refers to as "regurgitation" (i.e., what the New York Times demonstrated in its complaint) are both limited "bugs" in the model that OpenAI wants to eliminate. *Id.* During the editing process of this Article, OpenAI filed a motion to dismiss the N.Y. Times' Complaint, formally stating these same arguments. *See, e.g.*, Wes Davis, *OpenAI Claims The Times Cheated To Get ChatGPT To Regurgitate Articles*,

their models is fair use.[152] To date, however, no actual fair use determinations have been reached in any of the pending litigation.

As discussed in Section II.B.2, a fair use determination involves evaluating four factors against the purpose of "promot[ing] the progress of Science**.**"[153] This analysis, however, is very fact-intensive and, therefore, it is difficult to establish much precedential value from any given fair use decision.[154] For example, Generative AI companies point to the decision in *Google Books* to support their argument that training Generative AI models is fair use.[155]

As discussed above, however, the findings in *Google Books* were based heavily on the technological measures Google put in

---

THE VERGE (Feb. 27, 2024, 4:21 PM), https://www.theverge.com/2024/2/27/24084843/openai-new-york-times-dismiss-copyright-claims-lawsuit [https://perma.cc/FCJ4-YRB5].

[152] Numerous Generative AI companies have provided comments in response to the Copyright Office's request for comment on Artificial Intelligence and copyright. *Notice: Artificial Intelligence and Copyright*, U.S. COPYRIGHT OFFICE (Aug. 30, 2023), https://www.regulations.gov/document/COLC-2023-0006-0001/comment [https://perma.cc/X5GW-JTWB]. To find links to the comments from major companies in the over ten thousand responses, the Author recommends this article from The Verge, which links to comments from Apple, Adobe, Google, Meta, and Microsoft, among others. Wes Davis, *AI companies Have All Kinds of Arguments Against Paying for Copyrighted Content*, THE VERGE (Nov. 4, 2023, 6:17 PM), https://www.theverge.com/2023/11/4/23946353/generative-ai-copyright-training-data-openai-microsoft-google-meta-stabilityai [https://perma.cc/M9UP-MEPC].

[153] Authors Guild v. Google, Inc., 804 F.3d 202, 212 (2d Cir. 2015) (quoting U.S. CONST., art. I, § 8, cl. 8).

[154] Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 577 (1994) (explaining that fair use analysis is "not to be simplified with bright-line rules, for the statute, like the doctrine it recognizes, calls for case-by-case analysis.").

[155] *See, e.g.*, Ashley Belanger, *OpenAI Disputes Authors' Claims That Every ChatGPT Response is a Derivative Work*, ARS TECHNICA (Aug. 30, 2023, 1:31 PM), https://arstechnica.com/tech-policy/2023/08/openai-disputes-authors-claims-that-every-chatgpt-response-is-a-derivative-work [https://perma.cc/4W9J-FXXJ]. The Author notes that some also point to the Supreme Court's decision in *Google v. Oracle*, finding fair use in API copying as support for unlicensed use of copyrighted materials in Generative AI training. Google LLC v. Oracle, Inc., 141 S. Ct. 1183 (2021). The Author falls into the camp of those who believe this reliance is misplaced because *Google v. Oracle* is a very narrow opinion related to primarily functional computer code.

place in its search and snippet products to prevent those offerings from being a substitute for the original works.[156] The Second Circuit even pointed out that the fair use analysis could swing differently if Google had been "providing text in a manner that offers a competing substitute" of plaintiffs' works.[157]

If the New York Times can prove its factual allegations, it will demonstrate that exact type of fair-use-analysis-shifting substitution. For example, as part of the fourth fair use factor, the New York Times may be able to demonstrate that ChatGPT is serving as a complete substitute for the New York Times' "Wirecutter" product recommendation service.[158] Similarly, using the first factor standard set forth in *Andy Warhol Foundation v. Goldsmith*,[159] the New York Times may be able to show that the use of ChatGPT is a commercial purpose that leans against a finding of fair use[160] and that OpenAI's ChatGPT could fully reproduce a paywalled New York Times article and alleviate the need for going to the New York Times at all.[161]

The difference in facts between *Google Books* and the *New York Times* complaint demonstrates why fair use is such a difficult concept to rely on to get definitive (or at least precedential) answers

---

[156] *Authors Guild v. Google*, 804 F.3d at 216–17.

[157] *Id.* at 218.

[158] Complaint at 48–55, N.Y. Times Co. v. Microsoft Corp., No. 1:23-cv-11195 (S.D.N.Y. Dec. 27, 2023).

[159] Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith, 598 U.S. 508 (2023).

[160] *Id.* at 532–33 ("If an original work and a secondary use share the same or highly similar purposes, and the secondary use is of a commercial nature, the first factor is likely to weigh against fair use, absent some other justification for copying.").

[161] *See, e.g.*, Complaint at 29–46, *N.Y. Times Co.*, No. 1:23-cv-11195. As of the date of this Article, no lawsuits have been filed regarding Generative AI's ability to reproduce copyrighted imagery as discussed *supra* Section III.A.2. Though the fair use analysis would be different regarding reproduction of a still from a movie versus an entire textual work, the Author submits that reproduction in the outputs, combined with the many reproductions made during training, still set forth a strong copyright infringement claim. Detailed speculation about other legal theories that could be brought (for example, in the form of Lanham Act violations or state unfair competition) is beyond the scope of this Article.

regarding the unlicensed use of copyrighted materials in the training of Generative AI.[162] Every situation and every model could lead to a different result.[163] That level of uncertainty is untenable with the amount of money involved in creating these Generative AI models.[164]

---

[162] The Author notes that there are some who argue that copies made in training Generative AI models are nothing more than "intermediate copies" that have been held to be a fair use in previous cases. *See, e.g.*, Sega Enters. Ltd. v. Accolade, Inc., 977 F.2d 1510, 1527–28 (9th Cir. 1992), *amended* (Jan. 6, 1993) ("[W]here disassembly is the only way to gain access to the ideas and functional elements embodied in a copyrighted computer program and where there is a legitimate reason for seeking such access, disassembly is a fair use."). Though an intellectually interesting argument, the Author submits that the allegations presented in the New York Times Complaint and the image examples shown in Section III.A.2 suggest there is more than a temporary, transient intermediate copy of the data being made as a part of Generative AI training. In fact, the Author agrees with Cornell University Law School professor James Grimmelmann, who suggests that that argument might not "even be true on the facts." Ivan Moreno, *AI Models' Link To Nonprofit Data Raises Fair Use Question*, LAW360 (Jan. 18, 2024, 10:29 PM), https://www.law360.com/ip/articles/1787565 [https://perma.cc/TTN4-PG5K].

[163] Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 577 (1994) (stating that fair use analysis is "not to be simplified with bright-line rules, for the statute, like the doctrine it recognizes, calls for case-by-case analysis").

[164] Though this Article focuses on the application of copyright law to Generative AI, the Author recognizes that this interplay also has policy implications. There are conflicting positions at the policy level of whether companies should pay licensing fees to copyright holders for use their data, as some politicians and media companies believe is the appropriate path forward. Brian Merchant, *The AI Industry Has a Battle-Tested Plan to Keep Using Our Content Without Paying For It*, L.A. TIMES (Jan. 12, 2024), https://www.latimes.com/business/technology/story/2024-01-12/column-copyright-is-the-biggest-threat-to-the-ai-industry-but-its-not-going-down-without-a-fight [https://perma.cc/LS3X-SF4Z]. Apple, for example, is believed to be taking this approach regarding training its models. Jay Peters, *Apple Reportedly Wants to Use the News to Help Train its AI Models*, THE VERGE (Dec. 22, 2023, 4:32 PM), https://www.theverge.com/2023/12/22/24012730/apple-ai-models-news-publishers [https://perma.cc/L72D-HERN]. Reddit also recently announced it has signed a lucrative licensing deal with a large Generative AI company (though that company was not named). *Reddit Signs Content Licensing Deal With Al Company Ahead of IPO, Bloomberg Reports*, REUTERS (Feb. 16, 2024, 6:42 PM), https://www.reuters.com/technology/reddit-signs-content-licensing-

B. The Copyright Office Denies Applications for Generative AI Works

As discussed previously in Section II.A, a work must be original and created by a human author to be eligible for copyright protection.[165] The Copyright Office has faithfully applied that standard and denied registration to works partially or wholly generated by Generative AI.[166] The Copyright Office has also issued guidance regarding registering works containing AI.[167] Specifically,

---

deal-with-ai-company-ahead-ipo-bloomberg-reports-2024-02-16/ [https://perma.cc/S586-E4ZT]. OpenAI does pay to license some content for its training. Matt O'Brien, *ChatGPT-maker OpenAI Signs Deal With AP to License News Stories*, ASSOCIATED PRESS (July 13, 2023), https://apnews.com/article/openai-chatgpt-associated-press-ap-f86f84c5bcc2f3b98074b38521f5f75a [https://perma.cc/VCK6-SYML]. However, reports have come out that state that what OpenAI is willing to offer to license data does not align with OpenAI's high valuation. *Id. Compare* David, *supra* note 112, *with Microsoft-backed OpenAI Valued at Over $80 Billion in Latest Deal*, COMPETITION POL'Y INT'L (Feb. 18, 2024), https://www.pymnts.com/cpi_posts/microsoft-backed-openai-valued-at-over-80-billion-in-latest-deal [https://perma.cc/C958-FV9T]. Others argue, however, that AI companies "would go broke" if required to pay copyright royalties or licensing fees. Brian Merchant, *supra* note 164. Even still, others say that payment of licensing fees would restrict Generative AI to only the largest companies with the deepest pockets. Megan Morone, *Copyright Law is AI's 2024 Battlefield*, AXIOS (Jan. 2, 2024), https://www.axios.com/2024/01/02/copyright-law-violation-artificial-intelligence-courts [https://perma.cc/T8SU-Z9UP].

[165] *See, e.g.*, Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340 (1991); *see also* related discussion *infra* Section II.A.

[166] *See* discussion *infra* Section IV.B. The Author notes that this Article focuses on copyright jurisprudence in the United States and recognizes that other countries, including China, have begun granting intellectual property protection to AI-generated works. *See, e.g.*, Dennis Crouch, *Artificial Intelligence and Copyright in China: Lessons from a Recent Court Case*, PATENTLY-O (Dec. 28, 2023), https://patentlyo.com/patent/2023/12/artificial-intelligence-copyright.html [https://perma.cc/NT6U-7Z4J]. The Author also notes that the copyrightability standard differs from the standard used in determining patentability. The United States Patent and Trademark Office recently issued guidance related to the use of Generative AI in creating inventions and that use's impact on patentability. *See* UNITED STATES PATENT AND TRADEMARK OFFICE, INVENTORSHIP GUIDANCE FOR AI-ASSISTED INVENTIONS (2024).

[167] Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence, 37 C.F.R. § 202 (2023). In addition to existing

the Copyright Office requires that registrants specify what parts of their submissions were generated by AI instead of by a human author.[168] This information allows the Copyright Office to follow its guidelines regarding human authorship, even as it relates to only portions of works.

As a foundation for the discussion of the decisions related to works made by Generative AI, the Copyright Office's Guidance states:

> If a work's traditional elements of authorship were produced by a machine, the work lacks human authorship and the Office will not register it. For example, when an AI technology receives solely a prompt from a human and produces complex written, visual, or musical works in response, the "traditional elements of authorship" are determined and executed by the technology—not the human user.
>
> 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. Instead, these prompts function more like instructions to a commissioned artist—they identify what the prompter wishes to have depicted, but the machine determines how those instructions are implemented in its output.[169]

The decisions below will show how this guidance is applied in practice. Specifically, the Copyright Office has denied registration

---

guidance, the Copyright Office has said it plans to put out three reports this year setting forth its position on copyright law and Generative AI. Cecilia Kang, *The Sleepy Copyright Office in the Middle of a High-Stakes Clash Over A.I.*, N.Y. TIMES (Jan. 26, 2024), https://www.nytimes.com/2024/01/25/technology/ai-copyright-office-law.html [https://perma.cc/JUR4-HJ59]. Though these reports will not be binding on courts, they will likely be highly influential with judges making decisions, as well as with lawmakers and regulators. As explained above, *supra* note 152, the Copyright Office has asked for comment regarding Generative AI to help inform its opinions. *Notice: Artificial Intelligence and Copyright*, *supra* note 152. Of note, President Joe Biden's October 20, 2023 Executive Order on Artificial Intelligence was silent on the issue of Generative AI. *See Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*, WHITE HOUSE (Oct. 30, 2023), https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence [https://perma.cc/3AQ6-5JCS].

[168] 37 C.F.R. § 202.

[169] *Id.*

to: (1) works that were created solely by Generative AI;[170] (2) parts of works created by Generative AI, even when a human created the remaining parts of the work;[171] and (3) works where a human author used Generative AI tools that handle the creative decisions associated with creating a work, even if a human instructed the Generative AI tool more broadly.[172] These decisions are consistent with existing copyright law, and the only one of these decisions that has been appealed to the federal courts has been upheld.[173]

1. Works Created Solely by Generative AI

Stephen Thaler used a computer system he called the "Creativity Machine" to generate a piece of art entitled *A Recent Entrance to Paradise*.[174] Thaler tried to register the copyright in the work, claiming the machine as the author and himself as the registrant as a "work for hire."[175] The Copyright Office denied the application on the grounds the work failed to satisfy the human authorship requirement, and Thaler appealed.[176]

The U.S. District Court for the District of Columbia's decision denying Thaler's appeal and upholding the registration refusal contained an excellent summary of the human authorship requirement. Writing for the court, Judge Beryl Howell explained that "[t]he 1976 Act's authorship requirement as presumptively being human rests on centuries of settled understanding."[177] In support of this statement, she cited and discussed many of the same cases discussed above in Section II, including *Burrow-Giles*,[178] *Naruto*,[179] and *Kelly*.[180] Of particular note, Judge Howell recognized that there remained an open question as to "how much human input

---

[170] *See* discussion *infra* Section IV.B.1.
[171] *See* discussion *infra* Section IV.B.2.
[172] *See* discussion *infra* Section IV.B.3.
[173] Thaler v. Perlmutter, No. 22-1564, 2023 WL 5333236 (D.D.C. Aug. 18, 2023).
[174] *Id.* at *1.
[175] *Id. See also* NIMMER & NIMMER, *supra* note 4, § 1.06.
[176] *Thaler*, 2023 WL 5333236, at *1.
[177] *Id.* at *4.
[178] Burrow-Giles Lithographic Co. v. Sarony, 111 U.S. 53 (1884).
[179] Naruto v. Slater, 888 F.3d 418, 420 (9th Cir. 2018).
[180] Kelley v. Chi. Park Dist., 635 F.3d 290, 291 (7th Cir. 2011).

is necessary to qualify the user of an AI system as an 'author' of a generated work,"[181] but that the facts in *Thaler* were not so complex as to require that question to be addressed.[182] The facts of *Thaler*, as shown by the administrative record, were that Generative AI made a work without the involvement of a human.[183] That type of work is not eligible for copyright protection.[184]

2. Portions of a Larger Work Created Solely by Generative AI

Kris Kashtanova applied for copyright registration for the graphic novel *Zarya of the Dawn*,[185] which included original text and layouts created by Kashtanova and artwork made by Midjourney, a Generative AI model.[186] Upon re-reviewing the application, the Copyright Office indicated that the text and layouts created by Kashtanova were eligible for protection, but the images created by Midjourney were not.[187]

---

[181] *Thaler*, 2023 WL 5333236, at *6.

[182] *Id.*

[183] *Id.*

[184] *Id.* at *6–7. Though not necessary to understand this case, the Author provides n.2 from the *Thaler* opinion in full for the reader's amusement:

> The issue of whether non-human sentient beings may be covered by "person" in the Copyright Act is only "fun conjecture for academics," Justin Hughes, *Restating Copyright Law's Originality Requirement*, 44 COLUM. J.L. & ARTS 383, 408–09 (2021), though useful in illuminating the purposes and limits of copyright protection as AI is increasingly employed. Nonetheless, delving into this debate is an unnecessary detour since "[t]he day sentient refugees from some intergalactic war arrive on Earth and are granted asylum in Iceland, copyright law will be the least of our problems."

Id. at *4 n.2.

[185] Letter from Robert J. Kasunic, U.S. Copyright Off., to Van Lindberg, Taylor English Duma LLP (Feb. 21, 2023) (on file with Library of Congress) (regarding Zarya of the Dawn, registration # VAu001480196).

[186] *Id.*

[187] *Id.* Though not necessary to understand the legal aspects of the decision, the procedural history of this issue is interesting. Kashtanova originally submitted the work without indicating that that imagery in the graphic novel was AI generated. *Id.* at 3, 4. When Kashtanova posted on social media about a registration being granted related to an AI created work, the Copyright Office *sua sponte* re-evaluated the application on the basis that it lacked information during its initial review. *Id.* After a back-and-forth exchange with Kashtanova's

The Copyright Office's letter explaining its decision detailed how Midjourney worked, including how Midjourney provided the user with four different images based on a given user text prompt.[188] Based on this technical description, the Copyright Office explained that Midjourney was not "a tool that Ms. Kashtanova controlled and guided to reach [the] desired image" but, instead, a machine that "generates images in an unpredictable way."[189] Stated another way, the user's prompt of the Generative AI may "influence" how the AI tool generates an image, but it does not dictate a specific result. Consequently, the images generated by Midjourney were not like the images created by a camera, as discussed in *Burrow-Giles* (where the photographer had extensive creative input in the photographs, such as in the posing of the subject and arrangement of the scene related to the light)[190] because the machine, not the author, made all the creative decisions when using Midjourney.[191]

When viewed in conjunction with *Thaler*, this decision makes it clear that the Copyright Office is following its guidance[192] and does not view Generative AI works generated by textual prompting as eligible for copyright protection, even if those Generative AI works are included as part of a larger work.[193]

---

attorney, the Copyright Office modified the registration to exclude the AI generated images as discussed in this Section. *Id. See also* Benj Edwards, *Adventures in 21st Century Copyright*, ARS TECHNICA (Feb. 23, 2023, 1:19 PM), https://arstechnica.com/information-technology/2023/02/us-copyright-office-withdraws-copyright-for-ai-generated-comic-artwork [https://perma.cc/TA2D-Q9SS]. *See also* PATRY, *supra* note 15, § 3:19.

[188] Letter from Robert J. Kasunic to Van Lindberg, *supra* note 184, at 6–8.

[189] *Id.* at 9. The letter again addresses and dismisses the suggestion that providing textual prompts that result in the generated image satisfies the human authorship requirement on pages 9 and 10.

[190] Burrow-Giles Lithographic Co. v. Sarony, 111 U.S. 53, 53–54 (1884).

[191] Letter from Robert J. Kasunic to Van Lindberg, *supra* note 184, at 9. The Author notes this line of reasoning is similar to that used in denying copyright protection to a garden as discussed above in Section I.A.2. Kelley v. Chi. Park Dist., 635 F.3d 290, 291 (7th Cir. 2011).

[192] Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence, 37 C.F.R. § 202 (2023).

[193] *See also* Letter from Suzanne V. Wilson, U.S. Copyright Off., to Tamara Pester, Tamara S. Pester, LLC (Sept. 5, 2023) (on file with the Library of

3. Works That Include Both Human and Generative AI Contributions

The *Thaler* and *Zarya of the Dawn* decisions raise the question of what level of human involvement is required in a work's creation for the work to satisfy the human authorship requirement for copyrightability. Specifically, *Zarya* suggests that for Generative AI models functioning similarly to widely used image models like Midjourney and Dall-E, additional input beyond textual prompts is required.

In late 2023, Ankit Sahni tested the boundaries of human and Generative AI interaction by using a custom-built AI system known as RAGHAV Artificial Intelligence Painting App ("RAGHAV") to modify a photograph he had previously taken and make it resemble a painting in the style of Vincent van Gogh's *The Starry Night.* He titled this new work "SURYAST."[194] In his copyright application for SURYAST, Sahni listed himself as the author of "photograph, 2-D artwork" and RAGHAV as the author of "2-D artwork."[195]

In its decision denying protection to the work, the Copyright Office explained the differences between RAGHAV and more traditional photography editing tools such as Adobe Photoshop.[196] Specifically, RAGHAV's AI model itself "predicts stylizations for paintings and textures never previously observed," not the user.[197]

Congress) (regarding second request for reconsideration for refusal to register Théâtre D'opéra Spatial) (stating that the applicant's "sole contribution to the Midjourney Image was inputting the text prompt that produced it" because that means the "traditional elements of authorship" are determined by the machine, not the user).

[194] Letter from Suzanne V. Wilson, U.S. Copyright Off., to Alex P. Garens, Day Pitney, LLP (Dec. 11, 2023) (on file with the Library of Congress) (regarding second request for reconsideration for refusal to register SURYAST).

[195] *Id.* at 2.

[196] *Id.* at 8. The Author notes that as of early 2024, the Generative AI adjacent topic of computational photography (in which a camera creates a "photograph" from many different scenes and captures despite a single click of the shutter) has not been addressed by the Copyright Office. *See, e.g.*, Jay Peters, *The Pixel 8 and the What-is-a-Photo Apocalypse*, THE VERGE (Oct. 7, 2023, 8:30 AM), https://www.theverge.com/2023/10/7/23906753/google-pixel-8-pro-photo-editing-tools-ai [https://perma.cc/432U-QVUG].

[197] Letter from Suzanne V. Wilson to Alex P. Garens, *supra* note 193, at 8.

Sahni's input was limited to selecting the style and then deciding on the "strength" of that style by choosing a single number.[198] The Copyright Office found Sahni's contribution to be *de minimis* and not the type of contribution that constituted authorship.[199] It also rejected Sahni's argument that selecting the style to be applied in SURYAST satisfied the human authorship requirement.[200] The Copyright Office viewed his selection as developing an unprotectable idea for the work, not creating the protectable expression.[201]

When viewed in context with the *Thaler* and *Zarya of the Dawn* decisions, the *SURYAST* opinion suggests that for a work to be eligible for copyright protection, there must be human creative input that can be directly tied to the expression in the work.[202] Simply selecting from available options like image styles and letting Generative AI do the rest will not satisfy that requirement. Unfortunately, being unable to create protectible works when using Generative AI in this manner strips AI tools of some of their benefits, including potential increases in efficiency.

C. Consequences of Denying Works Created with Generative AI Copyright Protection

The decisions examining whether Generative AI works can receive copyright protection are interesting from an intellectual

---

[198] *Id.*

[199] *Id.* The Author notes that the Copyright Office's decision relates to the total work SURYAST and does not discuss the copyrightability of the underlying photograph. The Author submits that had Sahni applied to register the photograph on its own, the Copyright Office likely would have approved that application.

[200] *Id.* The Author notes that n.5 of the Letter discusses the lack of information about the creation of RAGHAV in the review record so the Copyright Office did not evaluate that process in reaching its decision. To date, there has not been a test case in which the author both created the Generative AI model and then used the model to create a work.

[201] *Id.*

[202] The Author is aware that this sentence may seem obvious when viewed against the textual language of the Copyright Act, but the expanding corpus of decisions trying to test it suggests it may not be as obvious as it appears. Similarly, the policy issues discussed above may ultimately override it. *See supra* text accompanying note 165.

perspective, but they also have practical implications, especially in the software development space. For example, it has become apparent that Generative AI is quite good at writing source code.[203] Intuitively, this makes sense because the rules (or grammar) of programming languages are much simpler than those of spoken languages.[204] Further, once an AI tool generates code, a programmer can easily run it against a compiler to ensure it works. This is a trivial interaction compared to reviewing text to ensure it is grammatically sound and has correct factual information.[205] For this reason, tens of thousands of companies are using Generative AI tools to generate source code.[206]

If all of this code created by Generative AI is ineligible for copyright protection,[207] however, that would result in new limitations on how software companies can protect one of their primary assets.[208] Currently, the use of computer software without a license or outside the scope of a license forms the basis for a

---

[203] Willison, *supra* note 81.

[204] *Id.*

[205] *Id.*

[206] *See, e.g.*, Heather Whitney, *Court Says No Human Author, No Copyright (but Human Authorship of GenAI Outputs Remains Uncertain) (Guest Blog Post*), TECH. & MKTG. L. BLOG (Aug. 22, 2023), https://blog.ericgoldman.org/archives/2023/08/court-says-no-human-author-no-copyright-but-human-authorship-of-genai-outputs-remains-uncertain-guest-blog-post.htm [https://perma.cc/J4R8-4CNX] ("Microsoft's CEO, Satya Nadella, said last month that *27,000 companies* are paying for a GitHub Copilot enterprise license.") (emphasis added).

[207] There are other legal issues related to Generative AI that are tangentially related to copyright law. For example, if Generative AI memorization creates an image for an end-user that is substantially similar to a copyrighted work, how is liability between the model and the end user determined under copyright's strict liability regime? Though it is traditionally understood that plaintiffs like to go after parties with deep pockets, the Author submits that some copyright plaintiffs prefer a "quantity over quality" strategy regarding infringement, especially in the realm of photographs. *See also* Megan Morrone, *Copyright Law Will Shape How We Use Generative Al*, AXIOS (Dec. 1, 2023), https://www.axios.com/2023/12/01/ai-copyright-law-biden-executive-order [https://perma.cc/VXC4-UNN3] (quoting Sean O'Brien (founder of the Yale Privacy Lab) that "[t]here will be software ecosystems polluted with proprietary code that will be the subject of cease- and-desist claims by enterprising firms").

[208] Tom LaPerle, Apple, Inc., Comment Letter on U.S. Copyright Office Artificial Intelligence and Copyright Study (Oct. 30, 2023).

copyright infringement claim.[209] If the code is not eligible for protection, however, that makes copyright infringement an unviable claim.

The lack of copyright protection for such a key company asset could also present itself during software company mergers and acquisitions. If a company widely uses Generative AI to create its source code, it may be impossible for the company to warrant to its buyer that it has the right, title, and interest in the source code asset it is trying to sell. Whether in the litigation or mergers and acquisitions context, this potential issue will require extensive investigation into the tools used by developers and analysis of which parts of source code were generated by humans instead of Generative AI.

Until the issue of source code ownership is resolved, software companies will be forced to face the difficult decision of trading the efficiency associated with using Generative AI for the protectability of human-authored code. This decision becomes even more complicated in rapidly evolving industries where being more efficient and getting a product to market faster can be the difference between a software company succeeding or failing.[210]

V. THE PERILOUS ROAD TO PRIVATE COPYRIGHT ENFORCEMENT

As discussed above, applying existing copyright law to Generative AI is difficult. The fact-intensive analysis required for a fair use determination makes it ill-fit to wholly address the *prima facie* copyright infringement that Generative AI models may be engaged in if they are trained by the unlicensed use of copyrighted

---

[209] Vernor v. Autodesk, Inc., 621 F.3d 1102, 1111 (9th Cir. 2010) (explaining that a licensee is not the owner of a copy and not eligible for use of certain copyright defenses).

[210] The Author understands that it is unlikely a piece of software will be wholly generated by Generative AI, but the Generative AI question—at a minimum—opens the door to more careful examination of what part of the software is actually being reproduced as part of an unlicensed use. This will likely involve targeted discovery into the way the software was developed, including how Generative AI tools were used. It will likely also involve a careful examination of what parts of the software are loaded into RAM during execution to see if any human created code is actually being loaded and, therefore, reproduced.

materials. Meanwhile, the lack of protection for works or parts of works created by Generative AI forces companies to make tradeoffs between efficiency and protectability. This void opens the door for large players in the technology space to begin or, in some cases, expand their efforts at private copyright enforcement.[211]

The concept of private copyright enforcement is not new. Large content platforms have had ways to manage digital content on their systems for years.[212] One of the most widely known of these systems is Google's YouTube Content ID program.[213] Content ID is a digital fingerprinting system wherein videos uploaded to YouTube are compared against audio and video files content owners have registered with the system.[214] If a Content ID match is found,

---

[211] While this Article focuses on the interplay of Generative AI and copyright law, the Author notes that much of the public concern over Generative AI relates to the area of disinformation, including "deepfakes" (i.e., use of Generative AI to replicate someone's likeness into situations where the person was not actually involved). Although some of the examples of deepfakes are rather innocuous, such as a picture of Pope Francis in a puffer jacket, the possibility of more nefarious uses of the technology is clearly troubling. Simon Ellery, *Fake Photos of Pope Francis in a Puffer Jacket Go Viral, Highlighting the Power and Peril of AI*, CBS NEWS (Mar. 28, 2023, 11:39 AM), https://www.cbsnews.com/news/pope-francis-puffer-jacket-fake-photos-deepfake-power-peril-of-ai [https://perma.cc/KBW3-MPDE]. On the subject of deepfakes specifically, there seems to be a public consensus regarding passing a law to address this conduct, including to criminalize it in certain circumstances. *See, e.g.*, Pranshu Verma, *AI Fake Nudes Are Booming. It's Ruining Real Teens' Lives,* WASH. POST (Nov. 5, 2023, 3:00 AM), https://www.washingtonpost.com/technology/2023/11/05/ai-deepfake-porn-teens-women-impact [https://perma.cc/PHS3-RJXP]. As of early 2024, none of the bills have been passed. The FTC, however, has banned the use of Generative AI in robocalls in the wake of a robocall call campaign where a deepfake of President Joe Biden urged people not to vote. Brian Bushard, *AI-Generated Robocalls Banned After Troubling Deepfakes*, FORBES (Feb. 8, 2024, 11:14 AM), https://www.forbes.com/sites/brianbushard/2024/02/08/ai-generated-robocalls-banned-after-troubling-deepfakes [https://perma.cc/28PC-Z7FT].

[212] *See, e.g.*, *How Content ID Works,* YOUTUBE HELP, https://support.google.com/youtube/answer/2797370 [https://perma.cc/52NX-6B3H] (last visited Apr. 2, 2024). For clarity, this Section will refer to YouTube (the video hosting platform owned by Google) instead of Google (which many associate with online search instead of online video).

[213] *Id*.

[214] *Id.*

YouTube takes actions against the uploaded video, such as blocking the video from being viewed or running ads from the content owner alongside the video.[215] Closely related to the Content ID system is YouTube's Copyright Strike system.[216] Under this system, when a copyright owner submits a takedown request against a user's video, YouTube applies a "copyright strike" to the user's account.[217] Much like in a game of baseball, if a user gets three strikes, the user is out, and their account and videos are deleted from YouTube.[218]

One of the problems with this type of private copyright enforcement regime is that there is no way to establish precedent or have the third party's decisions appealed to a neutral tribunal. This failure is because YouTube's system is not governed by intellectual property concepts, like copyright law or the DMCA's notice and takedown procedure,[219] but by YouTube's terms of service as a platform owner, which gives YouTube broad discretion related to the content on that platform.[220] This means these platforms are not required to accept standard copyright infringement defenses and exceptions such as fair use and parody as part of their policies. The decisions regarding the content and "infringement" are solely up to

---

[215] *Id.*

[216] *Copyright Strike Basics*, YOUTUBE HELP, https://support.google.com/youtube/answer/2814000 [https://perma.cc/4GXV-8PJ9] (last visited Apr. 2, 2024).

[217] *Id.*

[218] *Id.*

[219] The Author notes that the use of a DMCA-like notice and takedown system related to Generative AI models themselves may not be possible because experts are currently unsure if it's even possible to make models forget learned information. *See, e.g.*, Alison Snyder, *Machine forgetting: How Difficult It Is To Get AI To Forge*t, AXIOS (Jan. 2, 2024), https://www.axios.com/2024/01/12/ai-forget-unlearn-data-privacy.

[220] *See, e.g.*, Prager Univ. v. Google LLC, 951 F.3d 991, 995 (9th Cir. 2020) (upholding district court dismissal of lawsuit related to demonetization of videos); *see also* Hall v. Twitter, Inc., No. 20-cv-536-SE, 2023 WL 3322952 (D.N.H. May 9, 2023) (dismissing case against a platform related to account suspension); Craft v. Musk, No. 23-cv-01644-JCS, 2023 WL 2918739 (N.D. Cal. April 12, 2023) (dismissing case against a platform related to account suspension); Hall v. Meta, Inc., No. 3:22-cv-03063-TLB-MEF, 2022 WL 18109625 (W.D. Ark. Dec. 14, 2022) (dismissing case against a platform related to account suspension).

these third parties and frequently involve determinations related to the rights of select content partners that make the platforms substantial amounts of money.

The privatization path is further complicated because large technology companies are poised to serve as both Generative AI providers and content access gatekeepers. Companies such as YouTube[221] and Spotify[222] are platforms where people upload and access content, but they have also announced the availability of Generative AI tools. This opens the door for these companies to apply their private copyright enforcement policies strictly against content created with third-party Generative AI tools while applying them less stringently or not at all for content created by the platform's own tools. Relatedly, YouTube has also announced that it will expand its Content ID system to address works created by Generative AI that "simulate an identifiable individual."[223] Though YouTube will evaluate things like parody and satire for some videos, there will be no exceptions for parody if the "identifiable individual" is associated with one of YouTube's content partners.[224] This situation where a content partner can essentially overrule copyright defenses when dealing with private copyright enforcers is not present when addressing copyright issues in the court system.[225]

---

[221] Sarah Perez, *YouTube Shorts to Gain a Generative Al Feature Called Dream Screen*, TECHCRUNCH (Sept. 21, 2023, 10:25 AM), https://techcrunch.com/2023/09/21/youtube-shorts-to-gain-a-generative-ai-feature-called-dream-screen [https://perma.cc/8VEY-T597].

[222] Amrita Khalid, *Spotify is Going to Clone Podcasters' Voices — and Translate Them to Other Languages*, THE VERGE (Sept. 25, 2023, 8:00 AM), https://www.theverge.com/2023/9/25/23888009/spotify-podcast-translation-voice-replication-open-ai [https://perma.cc/8VEY-T597].

[223] Nilay Patel & Mia Sato, *YouTube is Going to Start Cracking Down on AI Clones of Musicians*, THE VERGE (Nov. 14, 2023, 6:33 AM), https://www.theverge.com/2023/11/14/23959658/google-youtube-generative-ai-labels-music-copyright [https://perma.cc/2Q7W-PLU2]. *See also* Neal Mohan, *Our Principles for Partnering with the Music Industry on Al Technology*, INSIDE YOUTUBE (Aug. 21, 2023), https://blog.youtube/inside-youtube/partnering-with-the-music-industry-on-ai [https://perma.cc/G8HC-ZQ53].

[224] *Id.*

[225] The Author recognizes differences in resource quantity can result in judicial inequity but believes the above discussed situation far exceeds that

Similarly, content providers such as Getty Images ("Getty") are providing their own Generative AI models.[226] As discussed in Section IV.A.2, current popular image Generative AI models may reproduce copyrighted imagery. Getty may try to leverage this uncertainty to claim that the only way to ensure that a user does not run afoul of the images in Getty's database is to use Getty's Generative AI tool instead of other alternatives.[227]

The private copyright enforcement efforts related to AI are not limited to just black and white infringement. For example, Meta recently announced that it would label AI-generated images across its various platforms—Facebook, Instagram, and Threads.[228] Other large companies are also working to identify AI-generated works,

---

concept. The Author submits it would be more like a plaintiff being able to pay a (much) larger filing fee at the outset in order to say that the defendant cannot use the fair use defense, not the plaintiff just being able to spend more on legal teams or attempt to use superior resources to run up the defendants' costs.

[226] Marty Swant, *Startup Debuts new Al Models Trained on Getty Images and Other Content as Copyright Concerns Loom*, DIGIDAY (Sept. 8, 2023), https://digiday.com/media-buying/startup-debuts-new-ai-models-trained-on-getty-images-and-other-content-as-copyright-concerns-loom [https://perma.cc/9LKC-ADWV].

[227] The Author recognizes that large Generative AI models have made public statements that they will indemnify end users for use of their models. *See, e.g.*, Kyle Wiggers, *OpenAI Promises to Defend Business Customers Against Copyright Claims*, TECHCRUNCH (Nov. 6, 2023, 1:15 PM), https://techcrunch.com/2023/11/06/openai-promises-to-defend-business-customers-against-copyright-claims [https://perma.cc/9UW9-B2NU]. A full examination of the terms of the various indemnification provisions from the large Generative AI providers is beyond the scope of this Article, but commentators have classified the indemnification as narrow, including because they exclude those models that are fine-tuned with a user's data. *See, e.g.*, Camilla Hodgson, *AI Firms' Pledges to Defend Customers from IP Issues Have Real Limits*, ARS TECHNICA (Jan. 8, 2024, 10:05 AM) https://arstechnica.com/ ai/2024/01/ai-firms-pledges-to-defend-customers-from-ip-issues-have-real-limits [https://perma.cc/B4J8-NWPZ].

[228] Aaron P. Rubin, *Meta Announces AI Content Identification for Facebook, Instagram, and Threads*, LEXOLOGY (Feb. 6, 2024), https://www.lexology.com/library/detail.aspx?g=fea227f3-235a-4334-8355-e9c2ce33942e [https://perma.cc/2X4P-3J6H].

including Google,[229] Amazon,[230] and Valve (who operates the large online game store, Steam).[231] The impact of this identification, including how it might interact with large technology companies' algorithms for displaying information to a user, is unclear at this time.[232]

Software companies may also experiment with ways to create a *de facto* copyright regime through contractual provisions. This is because the standard methods of enforcement of a breach of a software license may no longer be possible if Generative AI-created source code is not protectible under copyright, as discussed above in Section IV.B.3. Setting aside the difficulty in recreating concepts such as works made for hire that have such extensive histories, the contract-based enforcement mechanisms would only apply to the parties to the contract. There is no simple way to extend these concepts to a third party and use them to defend against a third party using the software.

The complication regarding copyright ownership also extends beyond software companies to areas like advertising, marketing, and design, where companies are hired to design assets (like logos) that have traditionally been the subject of copyright protection. These works for hire would not be eligible for such protection if Generative AI tools were used to create them. Absent changes to copyright, it may be simpler for companies to require artists and designers to

---

[229] Ina Fried, *Google Debuts a New Way to Watermark AI-generated Images*, AXIOS (Aug. 29, 2023), https://www.axios.com/2023/08/29/google-watermark-ai-generated-images [https://perma.cc/H3YA-VJZ7].

[230] Ella Creamer, *Self-publishers Must Declare if Content Sold on Amazon's Site is AI-generated*, THE GUARDIAN, (Sept. 11, 2023, 1:08 PM), https://www.theguardian.com/books/2023/sep/11/self-publishers-must-declare-if-content-sold-on-amazons-site-is-ai-generated [https://perma.cc/7TAU-KCSY].

[231] Kyle Orland, *Valve Now Allows the "Vast Majority" of AI-powered Games on Steam*, ARS TECHNICA (Jan. 10, 2024, 11:44 AM), https://arstechnica.com/gaming/2024/01/valve-most-games-made-with-ai-tools-are-now-welcome-on-steam [https://perma.cc/2LEA-HJ5P].

[232] Though this may have the benefit of limiting AI-generated disinformation, it also is dependent on how a platform applies this AI-generated tag to its algorithm. For example, Meta could refuse or limit the display information created by other AI models but may not apply the same level of scrutiny to items generated with its Emu Video and Emu Edit models.

warrant that they have not used Generative AI in creating their art instead of jumping through the contractual hoops that would be necessary to address this issue.[233]

As of early 2024, the full ramifications of a lack of copyright protection for Generative AI-made works are unknown. That, however, has not stopped private companies from trying to fill in the potential gaps while the courts and lawmakers work to address the questions. If the pace at which copyright law addresses Generative AI does not accelerate, these private copyright enforcement mechanisms—which lack some of the protections found in existing copyright jurisprudence—may become the default way creators try to enforce their rights in the Generative AI age.

VI. CONCLUSION

It is well established in copyright jurisprudence that works require originality and human authorship to be granted copyright protection.[234] For over one hundred and fifty years, courts have applied these requirements to emerging technologies and different types of proposed authorship to adapt copyright law to address new situations.[235] Traditional methods of addressing emerging issues in copyright law, such as fair use, are not equipped to handle the unique challenges presented by the operation of Generative AI and its

---

[233] The Author notes that the movie industry recently established a number of different boundaries regarding the use of AI in the areas of acting and screenwriting. *See, e.g.*, Ina Fried & Ryan Heath, *2. SAG-AFTRA Inks AI Deal for Voice Actors*, AXIOS AI+ (Axios, Arlington, Va.), Jan. 10, 2024, https://www.axios.com/newsletters/axios-ai-plus-57130c0e-ec9c-4590-b8d1-871006675f6a.html [https://perma.cc/2F7J-KS6Y]. *See also* Sara Fischer, *Hollywood Writers' Contract Deal Includes Historic AI Rules*, AXIOS (Sept. 26, 2023), https://www.axios.com/2023/09/27/ai-wga-hollywood-writers-contract [https://perma.cc/2V25-3VB9]. A full examination of the details of these agreements is beyond the scope of this Article. The Author notes, however, that both are examples of using contractual methods to address Generative AI instead of copyright law.

[234] *See, e.g.*, Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340 (1991); *see also* discussion *supra* Section I.A.

[235] *See, e.g.*, Burrow-Giles Lithographic Co. v. Sarony, 111 U.S. 53, 58 (1884); Naruto v. Slater, 888 F.3d 418, 420 (9th Cir. 2018); *see also* discussion *supra* Sections I.A.1, I.A.2, and I.B.1.

reliance on vast data sets for training.[236] This disconnect is already present in both the existing lawsuits between copyright holders and Generative AI companies and in the copyright registration process.[237] The potential lack of copyright protection for items like source code that are partially developed by Generative AI could have significant repercussions for industries that would benefit from leveraging Generative AI tools but may be hesitant to do so without changes to copyrightability requirements.[238]

Throughout its history, copyright law has been forced to account for new technologies while also trying "[*t*]*o promote the Progress of Science* . . . by securing for limited Times to Authors . . . the exclusive Right to their respective Writings."[239] The rapid growth in the development of Generative AI, however, poses a problem that existing copyright law is currently unsuited to solve. If something is not done to address Generative AI within the copyright ecosystem, the troubling reality is that copyright enforcement may be on a path to privatization.[240] The next few years will be critical in determining whether that occurs.

[236] *Compare* discussion *supra* Section II.B.2 *with* Section IV.A.3.
[237] *See* discussion *supra* Sections IV.A.1–IV.A.2 and IV.B.
[238] *See* discussion *supra* Section IV.C.
[239] U.S. CONST., art. I, § 8, cl. 8 (emphasis added).
[240] *See* discussion *supra* Section V.