

Developing a Coherent National Strategy: Artificial Intelligence (AI) and Copyright Law

Nikki Chavez Brown

I. INTRODUCTION

In 2023, half of senior information technology business executives surveyed worldwide considered artificial intelligence (AI) to be a technology of strategic importance that they would prioritize in 2024.¹ Between 2017 and 2020, funding of AI startups increased from \$18 billion to \$26 billion.² In 2023, seventy-three percent of U.S. companies have begun implementing AI technologies.³ The growing interest in AI is not limited to corporations. U.S. citizens have particularly shown an interest in generative AI. In fact, a 2023 survey of U.S. adults showed that respondents were most interested in generative AI content from members of the creative community such as artists and musicians, entertainment providers, and skilled amateurs and digital artists.⁴

As generative AI tools have become increasingly accessible to the general public, legal scholarship has examined the role that copyright law will play in AI innovation. This scholarship has largely focused on whether AI generated content is copyrightable, who should hold that copyright, and how to treat use of copyrighted works in training AI.⁵ However, with slow federal action to address AI, these discussions have largely focused on a hypothetical national strategy based on comparisons to foreign copyright laws and the general legal principles and philosophical frameworks underlying intellectual property law.⁶ Now, a hypothetical national strategy is no longer necessary.

¹ *The 2024 IT Outlook Report*, RACKSPACE TECHNOLOGY (Dec. 4, 2023).

² Bergur Thormundsson, *AI Startup Company Funding Worldwide 2020-2023*, by *Quarter*, STATISTA, (Dec. 14, 2023), <https://perma.cc/847F-TXGP>.

³ *2024 AI Business Predictions*, PWC, (Dec. 5, 2023), <https://perma.cc/4JFU-Z492>.

⁴ Bergur Thormundsson, *Appeal of Generative AI in Social Media in the U.S. 2023*, by *Content Creator Category*, STATISTA (Jan. 2023), <https://perma.cc/28V6-GN9U>.

⁵ See, e.g., Atilla Kasap, *Copyright and Creative Artificial Intelligence (AI) Systems: A Twenty-First Century Approach to Authorship of AI-Generated Works in the United States*, 19 WAKE FOREST J. BUS. & INTELL. PROP. L. 335 (Summer 2019); Victor M. Palace, *What if Artificial Intelligence Wrote This? Artificial Intelligence and Copyright Law*, 71 FLA. L. REV. 217 (Jan. 2019); Jessica L. Gillotte, *Copyright Infringement in AI-Generated Artworks*, 53 U.C. DAVIS L. REV. 2655 (June 2020).

⁶ See, e.g., Megan Svedman, *Artificial Creativity: A Case Against Copyright for AI-Created Visual Artwork*, 9 IP THEORY 1 (2020); Tzipi Zipper, *Mind Over Matter: Addressing Challenges of Computer-Generated Works Under Copyright Law*, 22 WAKE FOREST J. BUS. & INTELL. PROP. L. 129 (Winter 2022); Yudong Chen, *The Legality of Artificial Intelligence's Unauthorized Use of Copyrighted Materials Under China and U.S. Law*, 63 IDEA 241 (2023).

Congress, the White House, and various federal agencies have begun developing their own strategies to address AI.

In February 2020, the U.S. Copyright Office (USCO) began to publicly discuss the relationship between copyright law and AI.⁷ In a symposium co-sponsored by the USCO and the World Intellectual Property Organization (WIPO), panelists discussed the copyright implications of AI on both developers and end-users.⁸ Developers have largely been concerned with the treatment of copyrighted works in training AI systems. On the other hand, end-users have largely been concerned with the copyrightability of AI outputs. Moreover, members of the creative community have voiced concerns over both potential copyright infringement and the copyrightability of works utilizing AI technologies.⁹ Despite the increasing questions over generative AI in the creative community, the USCO did not issue a public statement on AI or host another AI-related event until October 2021.¹⁰

Finally, in March of 2023, the USCO officially launched an AI initiative “to examine the copyright law and policy issues raised by [AI], including the scope of copyright in works generated using AI tools and the use of copyrighted materials in AI training.”¹¹ The same day, the Office issued new registration guidance, clarifying that (1) applicants have a duty to disclose AI-generated content in works submitted for registration, and (2) a work whose “traditional elements of authorship were produced by a machine” are not eligible for registration, because it “lacks human authorship.”¹²

The March 2023 registration guidance answered some questions for end-users by clarifying that AI-generated works are ineligible for copyright registration. However, it left many unanswered questions regarding the level of creative control needed for a work to be eligible for registration and the use of copyrighted materials in training AI systems. Throughout the Spring of 2023, the Office hosted public

⁷ Symposium, *Copyright in the Age of Artificial Intelligence* (U.S. Copyright Off. & World Intell. Prop. Org. Feb. 5, 2020).

⁸ *Id.*

⁹ Nora Scheland, #ICYMI: The Copyright Office Hears from Stakeholders on Important Issues with AI and Copyright, LIBR. OF CONG. BLOGS (July 18, 2023), <https://perma.cc/P828-6PHN>.

¹⁰ Whitney Levandusky, *Artificial Intelligence: The Copyright Connection*, LIBR. OF CONG. BLOGS (Oct. 13, 2021), <https://perma.cc/M7X3-4JLA>; Conference, *Copyright Law and Machine Learning for AI: Where are We and Where are We Going?* (U.S. Copyright Off. & USPTO Oct. 26, 2021).

¹¹ *Copyright Office Launches New Artificial Intelligence Initiative*, COPYRIGHT.GOV: NEWSNET (Mar. 16, 2023), <https://perma.cc/79KJ-RGAT>.

¹² Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence, 88 Fed. Reg. 16190, 16192-93 (Mar. 16, 2023) (to be codified at 37 C.F.R. pt. 202).

listening sessions on the goals and concerns regarding generative AI in creative fields.¹³ These listening sessions culminated in the Office issuing a notice of inquiry and request for comment (NOI) in August 2023.¹⁴ In response, a number of AI industry leaders and members of the creative community submitted comments sharing their views on generative AI and the role that copyright law should play in its regulation.¹⁵ The Office has announced that it will be issuing a Report in several Parts to analyze these issues.¹⁶

While the Copyright Office continues to examine these questions and concerns, other sectors of the U.S. federal government and industries have begun adopting generative AI strategies and best practices. Scholarship is plentiful on what the “ideal” copyright regulations should be for generative AI.¹⁷ Yet, as interested parties have begun developing various strategies for the adoption of generative AI, it is important to examine how copyright law may impair or facilitate a coherent national approach to AI. This Article first examines the approaches, opinions, and concerns of the federal government and interested parties in regard to generative AI. Then, this Article examines the ways in which existing and future copyright regulations may impair or benefit these approaches to generative AI.

II. ANALYSIS

A. The Federal Government’s Approach to Generative AI

Since at least 2020, Congress, the White House, and various federal agencies have been evaluating the potential benefits and detriments of AI innovation. According to Congress, the overarching federal strategy for AI innovation is to achieve international dominance in the global AI market while safeguarding the rights and privacy of Americans.¹⁸ The White House has echoed this sentiment through actions focused on harnessing the benefits and mitigating the risks of AI.¹⁹ Intellectual property rights and regulations could significantly hamper these federal strategies. Yet, intellectual property considerations have been noticeably absent from these strategies, with most issued statements and guidance simply stating that expert

¹³ COPYRIGHT.GOV, *supra* note 11.

¹⁴ Artificial Intelligence and Copyright, 88 Fed. Reg. 59942 (USCO, Aug. 30, 2023).

¹⁵ *See infra* Part II.B.

¹⁶ On July 31, 2024, the Office published Part 1 of the Report, which calls for new federal legislation to address the use of AI to create unauthorized digital replicas or “deepfakes.” U.S. COPYRIGHT OFF., COPYRIGHT AND ARTIFICIAL INTELLIGENCE, PART 1: DIGITAL REPLICAS (July 31, 2024).

¹⁷ *See, e.g.*, Kasap *supra* note 5; Palace *supra* note 5; Gillotte *supra* note 5.

¹⁸ National Artificial Intelligence Initiative Act of 2020, 15 U.S.C. § 9411(a).

¹⁹ *Administrative Actions on AI*, AI.GOV, <https://perma.cc/ENX8-Z3NX>.

advice is needed in how to manage intellectual property rights. By outlining the federal strategies to AI, the following discussions provide an overview of the policy concerns to be considered in shaping a coherent national AI strategy.

1. Congress

To address AI, Congress has taken various approaches, including a task force, committee hearings, and proposed legislation. In 2020, Congress directed the National Science Foundation and the White House Office of Science and Technology Policy to create the National Artificial Intelligence Research Resource (NAIRR) Task Force.²⁰ The Task Force’s Final Report noted four goals: to spur innovation, to increase diversity of talent, to improve capacity, and to advance trustworthy AI.²¹

The Task Force proposed an “open research environment”²² supported by a national system of computational and data resources, testbeds, and software and testing tools.²³ Nevertheless, the Task Force recognized that “fostering an open research environment has tradeoffs with... protecting intellectual property rights.”²⁴ To address IP concerns, the Task Force suggested using licensing agreements to define “publication responsibilities, disposition of intellectual property arising out of the use of the data, ownership of derived datasets, and expectations for disposal of the data.”²⁵ However, the Task Force also noted that expert advice is needed on intellectual property management and agreements.²⁶

In addition to the task force, Congress held multiple hearings on AI risks and innovation in 2023 and early 2024.²⁷ Witnesses at these hearings have included federal agencies, scholars, and industry leaders.²⁸ Generally, the hearings have examined the

²⁰ NAT’L AI RSCH. RES. TASK FORCE, STRENGTHENING AND DEMOCRATIZING THE U.S. ARTIFICIAL INTELLIGENCE INNOVATION ECOSYSTEM, ii (2023).

²¹ *Id.* at v.

²² *Id.* at 26.

²³ *Id.* at iv–v.

²⁴ *Id.* at 26.

²⁵ *Id.* at 35.

²⁶ *Id.* at 13.

²⁷ Oma Seddiq & Elizabeth Kim, *AI Influencers Pound Capitol Hill Hallways to Shape Legislation*, BLOOMBERG LAW (Aug. 8, 2023), <https://perma.cc/KUU2-G3JL>; *Congressional Committees Continue AI Hearings in 2024*, AKIN (Jan. 10, 2024), <https://perma.cc/Z7M5-7NG2>.

²⁸ See *Hearing Wrap Up: Federal Government Use of Artificial Intelligence Poses Promise, Peril*, COMM. ON OVERSIGHT & ACCOUNTABILITY (Sept. 15, 2023), <https://perma.cc/14Z2-5DSQ>; *Oversight of A.I.: Legislating on Artificial Intelligence: Hearing Before the Subcomm. On Priv., Tech., & Law of the S. Comm. On the Judiciary*, 118th Cong. (2023), <https://perma.cc/5JQH-DZHK>.

current uses of AI and the ways in which Congress may address the potential harms of AI, such as privacy and security concerns.²⁹

Finally, a number of bills regarding AI technologies have been introduced in the 118th Congress.³⁰ These bills vary greatly, but they generally aim to impose restriction and regulations on AI systems to safeguard privacy, to implement research initiatives, and to address the fraud, lying and lack of transparency issues involved with AI.³¹ Many of these bills, especially since January 2024, have specifically focused on curbing “AI fraud” in which AI is used to impersonate the likeness and voice of individuals.³² Other notable bills have aimed to address issues such as bias in algorithmic systems, agency transparency when using automated systems for public interaction and critical decision-making, and coordinating international AI strategies.³³

These actions demonstrate that Congress is aware and, at least somewhat, focused on AI technologies. However, as of February 2024, Congress had passed no legislation, leaving the U.S. straggling behind the more than 30 countries that have passed some form of AI legislation.³⁴

2. The White House

In October of 2022, the White House Office of Science and Technology Policy (OSTP) issued the *Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People*, which outlines general principles for deploying AI systems in a way that recognizes civil rights and democratic values.³⁵ In developing the “Blueprint”, the OSTP conducted a year-long process of panel discussions, public listening sessions, meetings, formal requests for information, and public input from people throughout the United States, public servants across federal agencies, and

²⁹ See *Hearing Wrap Up*, *supra* note 28.

³⁰ *Artificial Intelligence Legislation Tracker*, BRENNAN CTR. FOR JUST. (Aug. 7, 2023, last updated Jan. 5, 2024), <https://perma.cc/WEN4-CHS8>. It is important to note that AI legislation was passed prior to the 118th Congress. That legislation was more broad than recent legislation, as it focused on initiatives to gather more information and funding provisions. See, e.g., National Artificial Intelligence Initiative Act of 2020, 15 U.S.C. § 9411; Artificial Intelligence Training for the Acquisition Workforce Act, 117th Cong., 136 Stat. 2238 (2022).

³¹ *Id.*

³² See, e.g., H.R. 7120, 118th Cong. (2024); H.R. 7123, 118th Cong. (2024); H.R. 6943, 118th Cong. (2024).

³³ S. 3478, 118th Cong. (2023); H.R. 6886, 118th Cong. (2023); H.R. 6425, 118th Cong. (2023).

³⁴ Bill Whyman, *AI Regulation is Coming—What is the Likely Outcome?*, CTR. FOR STRATEGIC & INT'L STUD. (Oct. 10, 2023), <https://perma.cc/5Z5L-WDEM>.

³⁵ OFF. OF SCI. & TECH. POL'Y, BLUEPRINT FOR AN AI BILL OF RIGHTS: MAKING AUTOMATED SYSTEMS WORK FOR THE AMERICAN PEOPLE 3 (2022).

members of the international community to examine the promises and potential harms of AI.³⁶

The final Blueprint outlined five principles: (1) safe and effective systems, (2) algorithmic discrimination protections, (3) data privacy, (4) notice and explanation, and (5) human alternatives, consideration, and fallback.³⁷ However, the Blueprint only vaguely addresses IP concerns as a potential limitation on the independent evaluation of and meaningful access to examine automated systems.³⁸

Moreover, in October of 2023, President Biden issued an Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence.³⁹ To “advance and govern the development and use of AI,” the Executive Order laid out a framework of guiding principles.⁴⁰ These principles generally focus on (1) improving AI security and safety; (2) promoting innovation, competition, and collaboration in order to harness the benefits of AI and establish the U.S.’s dominance in the AI market; (3) safeguarding jobs and consumers; and (4) protecting privacy and civil liberties.⁴¹

Again, these actions demonstrate that the Biden Administration is aware of and somewhat focused on AI technologies. However, these guiding principles have the potential to be significantly compromised if copyright regulations hamper access to the data needed to promote the innovation of safe AI systems.

3. The Department of Defense

The Department of Defense (DoD) has developed its own plan to implement AI systems as part of the National Defense Strategy. In Tenet 3 of its *Responsible Artificial Intelligence Strategy and Implementation Pathway*, the DoD explicitly states that it may seek AI vendors, including commercially available technologies.⁴² Similar to the Blueprint, Line of Effort (LOE) 3.2.5 of the DoD’s strategy and implementation pathway notes that the strategy will seek to “preserve privacy and civil liberties” to avoid “unintended bias in the design and development of AI capabilities that involve

³⁶ *Id.* at 4.

³⁷ *Id.* at 5–7.

³⁸ *Id.* at 20, 51.

³⁹ Exec.Order. No. 14110, Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, 88 Fed. Reg. 75191 (Oct. 30, 2023).

⁴⁰ *Id.*

⁴¹ *Id.* at 75191–93.

⁴² DOD RESPONSIBLE AI WORKING COUNCIL, U.S. DEP’T OF DEF. RESPONSIBLE A.I. STRATEGY AND IMPLEMENTATION PATHWAY, 25 (Jun. 2022).

the use of personal information.”⁴³ Moreover, in LOE 5.3.1, the DoD states that it will seek “interoperability with partners and allies with data, compute, and storage systems, software, and schema.”⁴⁴ As the DoD implements third party AI systems, it will be imperative that these systems are safe and effective. As such, burdens on the ability to use copyrighted materials in training AI systems may hamper these efforts.

B. Industry Approaches to AI

At the same time that the federal government has been developing its own strategies to AI innovation, industries have continued to develop and implement AI technologies. In particular, the Copyright Office’s NOI provided valuable insight into the inner-workings and opinions of AI developers and creative communities.⁴⁵

In short, the conflict between AI developers and the creative community is two-fold. First, AI developers argue that use of copyrighted works in AI training falls under the fair use doctrine of copyright law. Under the creators’ view, licensing agreements would be required before copyrighted works could be used in AI training. Second, AI developers argue that safeguards such as “opt-out” approaches and mechanisms to restrict potentially infringing outputs are sufficient to protect the rights of copyright owners.⁴⁶ Creators, on the other hand, argue for an “opt-in” approach and improved access to litigation and remedies for infringement.⁴⁷

1. AI Developers

In response to the Copyright Office’s NOI, a number of AI innovators submitted comments which outlined their visions, opinions, and strategies for AI.⁴⁸ The overarching view of AI developers has been that training AI models qualifies as a fair use under copyright law and that AI outputs have the potential to bring significant benefits to the creative community.⁴⁹ To support this view, AI developers provided insight into their training process and mechanisms used to safeguard the interests of copyright owners.

⁴³ *Id.* at 27.

⁴⁴ *Id.* at 30.

⁴⁵ *Artificial Intelligence and Copyright*, *supra* note 14.

⁴⁶ *See infra* Part II.B.1.

⁴⁷ Nat’l Writers Union, Comment Letter on Artificial Intelligence and Copyright 14-15 (Oct. 30, 2023), <https://perma.cc/Y4UH-ZUGL>.

⁴⁸ OpenAI, Comment Letter on Artificial Intelligence and Copyright 11 (Oct. 30, 2023), <https://perma.cc/H46C-WP8H>; Meta, Comment Letter on Artificial Intelligence and Copyright 10 (Oct. 30, 2023), <https://perma.cc/6M67-WLR6>.

⁴⁹ *See, e.g.*, OpenAI, *supra* note 48, at 2 (stating that generative AI tools may benefit copyright industries by “increase[ing] worker productivity, lower[ing] the costs of production, and stimulat[ing] creativity by making it easier to brainstorm, prototype, iterate, and share ideas”).

First, AI developers clarified that the AI models do not store copies of copyrighted works used in the training process. Instead, these models consist of parameters reflecting the “statistical relationship between different words in different scenarios,”⁵⁰ influenced by billions of diverse pieces of training data.⁵¹ Moreover, OpenAI, creator of ChatGPT, explained that it has employed measures to reduce verbatim repetition of training information, such as de-duplicating training data, working with copyright owners to identify internet sites reproducing their copyrighted works to exclude from training, and training ChatGPT to recognize and decline to respond to prompts seemingly aimed at reproducing significant portions of copyrighted works.⁵²

Additionally, both OpenAI and Meta outlined their stance that industry practices and existing copyright laws are sufficient to address potential infringing uses of AI tools. Industry leaders have implemented promising features to protect copyright owners through various opt-out features and training models to decline requests that could generate a potentially infringing output.⁵³ However, OpenAI also recognized that users seeking to generate infringing outputs may occasionally be able to evade its guardrails. If infringement occurs, OpenAI argues that the existing doctrines of secondary liability are sufficient to address the potential liability of an AI model creator or service provider.⁵⁴

2. Creative Communities

The creative community has begun utilizing generative AI in a myriad of ways. A recent estimate by Alina Valyaeva suggested that over 15 billion images have been generated using text-to-image AI models.⁵⁵ Creators have been able to expand their audiences by utilizing AI translation tools to translate podcasts and videos into various languages.⁵⁶ Advertisers have used OpenAI’s GPT-3 to create scripts for their ads.⁵⁷ Filmmakers have used AI tools to more efficiently create visual effects.⁵⁸ Software

⁵⁰ *Id.* at 6.

⁵¹ Meta, *supra* note 48, at 6.

⁵² OpenAI, *supra* note 48, at 7.

⁵³ *Id.*; Meta, *supra* note 48, at 7.

⁵⁴ OpenAI, *supra* note 48, at 14.

⁵⁵ Alina Valyaeva, *AI has Already Created as Many Images as Photographers Have Taken in 150 Years. Statistics for 2023*, EVERYPIXEL JOURNAL (Aug. 15, 2023), <https://perma.cc/C4WB-4V28>.

⁵⁶ OpenAI, *supra* note 48, at 2–3.

⁵⁷ *Id.*

⁵⁸ *Id.*

developers are utilizing AI tools to streamline coding tasks.⁵⁹ Nevertheless, as AI tools have proliferated, the creative community has been split on whether AI should be seen as a promising tool to improve efficiency or as an existential threat to creators.⁶⁰ Public awareness of the tension between corporations and creators boomed in the wake of the 2023 writer’s strike and Universal Music Group’s (UMG) decision not to renew its licensing agreement with social-media company TikTok in early 2024.⁶¹

Organizations like UMG have voiced a number of concerns over AI models. Notably, UMG argues that AI models violate IP rights by training on unlicensed, copyrighted material; reproducing, manipulating, and processing those copyrighted materials for financial gain without license or permission from copyright owners; generating outputs strikingly similar to copyrighted works; and “clon[ing] artists’ images, likenesses, and voices to create audio and visual works...disturbingly similar to the artists themselves.”⁶²

As such, creators have argued for a number of copyright and other regulations. First, the use of copyrighted materials in AI training should be considered an infringing use absent a licensing agreement.⁶³ Second, outputs that impersonate artists and their works should be considered infringing uses and a violation of unfair competition laws.⁶⁴ Third, labeling requirements should be imposed on AI-generated works to reduce misidentification and deception.⁶⁵ Fourth, creators should be notified of AI developers’ intent to use their works in training, regardless of whether they hold registered or common-law copyrights.⁶⁶

C. Creating a Coherent AI System

The creation of a coherent AI system must balance continued AI innovation with the intellectual property rights of creators, and the regulation of AI through copyright law has the potential to drastically shape the AI industry.

⁵⁹ *Id.*

⁶⁰ National Writers Union, *supra* note 47, at 5–6; *These Artists are Using AI as a Creative Partner. See How!*, WORKLIFE. (Feb. 22, 2023), <https://perma.cc/RM9P-VZBD>.

⁶¹ Los Angeles Times Staff, *Writers Strike: What Happened, How it Ended, and its Impact on Hollywood*, LA TIMES (Last Updated Oct. 19, 2023), <https://perma.cc/LCM6-TSYH>; Universal Music Group, *An Open Letter to the Artist and Songwriter Community—Why We Must Call Time Out on TikTok* (Jan. 30, 2024), <https://perma.cc/7QYL-MJCV>.

⁶² Universal Music Group, Comment Letter on Artificial Intelligence and Copyright 9 (Oct. 31, 2023), <https://perma.cc/4ZBJ-3S9Y>.

⁶³ *Id.* at 19–21.

⁶⁴ *Id.* at 22–23.

⁶⁵ National Writers Union, *supra* note 47, at 9–10.

⁶⁶ *Id.* at 12.

Both the federal government and AI developers appear to largely support an open-source system of AI innovation as a mechanism for improving the accessibility, safety, and efficacy of AI systems.⁶⁷ In fact, the innovation boom in generative AI has largely been driven by open-source libraries for building and training machine learning models. For example, open-source libraries like Common Crawl have been utilized in the training and research of AI models like OpenAI's GPT-3.⁶⁸

Open-source libraries are intended to make code available to the public to promote its development and refinement. As such, an open-source system is more likely to level the playing field between small developers and dominant firms with access to key assets like computing power, cloud storage, and data. Additionally, open-source systems, as envisioned by the NAIRR Task Force, may reduce barriers to participation in AI research and increase the diversity of AI researchers by increasing accessibility to a wider range of users and providing a platform for use in education and community-building activities.⁶⁹ Furthermore, the safety and efficacy of AI systems depends in part on access to diverse training datasets.⁷⁰ With the DoD seeking to increase implementation of AI through acquisition strategies,⁷¹ AI systems trained on limited datasets could pose significant national security challenges.

At the same time, broad access to AI can also increase the risk of copyright infringement and consumer deception if not regulated appropriately. The March 2023 guidance seemed to solidify that AI outputs are not eligible for copyright registration.⁷² As such, much of the debate has revolved around whether using copyrighted works in AI training datasets constitutes fair use.

The open-source movement is seemingly in direct contradiction to the general principles of copyright law. Affording authors of original works exclusive rights to

⁶⁷ Congress has expressed support for an open-source system through the NAIRR final report and bills that have been introduced. For example, the NAIRR final report proposes the creation of an “integrated portal” comprised of a “federated mix of computational and data resources.” NAT'L AI RSCH. RES. TASK FORCE, *supra* note 20, at v. In addition, bills such as H.R. 6881, 118th Congress (2023) advocate for increased transparency into the training data used in AI models. Moreover, President Biden's Executive Order calls for the promotion of “a fair, open, and competitive ecosystem and marketplace for AI” by addressing the risks of dominant firms' use of key assets and ensuring that “small developers and entrepreneurs can continue to drive innovation.” *See* EO 14110, *supra* note 39, at §2(b).

⁶⁸ *Mozilla Report: How Common Crawl's Data Infrastructure Shaped the Battle Royale over Generative AI*, MOZILLA (Feb. 6, 2024), <https://perma.cc/P8RW-LU8K>.

⁶⁹ *See* NAT'L AI RSCH. RES. TASK FORCE, *supra* note 20 at 9.

⁷⁰ Adam Zewe, *Can Machine-Learning Models Overcome Biased Datasets?*, MIT NEWS (Feb. 21, 2022), <https://perma.cc/9XZK-9FYW>.

⁷¹ *See* DoD RESPONSIBLE AI WORKING COUNCIL, *supra* note 42.

⁷² Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence, 88 Fed. Reg. 16190, 16192–93 (USCO Mar. 16, 2023) (to be codified at 37 C.F.R. 202).

reproduce, publicly display, and prepare derivatives of the copyrighted work has been traditionally viewed as a mechanism for encouraging creativity.⁷³ At the same time, however, copyright law has long recognized that granting copyright must be balanced against the public interest in the availability of literature, music, and other arts. As a result, the Copyright Act has limited the scope and reach of copyright owners' exclusive rights.⁷⁴

According to AI developers, AI training is a transformative use because it creates “a product with wholly different purposes, capabilities and uses.”⁷⁵ The training process creates AI models that store “statistical relationships between words, shapes, colors, textures, and concepts,”⁷⁶ rather than copies of the copyrighted works.⁷⁷ These statistical relationships, the argument goes, are unprotectable because copyright law protects the expression of facts and ideas rather than the facts and ideas themselves.⁷⁸ Under this theory, the purpose of utilizing copyrighted materials is to refine the algorithm rather than “to communicate through artistic expression, earn a living, and/or practice their craft.”⁷⁹ In this sense, training AI models has a different purpose than employing those models to generate expressive works.

Conversely, the creative community has argued that utilizing unlicensed copyrighted materials in AI training should be considered copyright infringement. The output of generative AI models, the theory goes, serve the same purpose as the copyrighted works used in training and thus, cannot be considered transformative.⁸⁰ Additionally, holding AI training as a fair use would impact the potential market by denying creators licensing income and creating competition between AI outputs and creators.⁸¹ Therefore, AI developers should be required to license copyrighted works

⁷³ *Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith*, 598 U.S. 508, 526 (2023).

⁷⁴ *Id.* at 526 (citing *Twentieth Century Music Corp. v. Aiken*, 422 U.S. 151, 156 (1975)); Copyright Act, 17 U.S.C. §§102, 107–122, 302–305.

⁷⁵ Van Lindberg, *Building and Using Generative Models Under U.S. Copyright Law*, 18 RUTGERS BUS. L. REV. 1, 38 (Spring 2023).

⁷⁶ OpenAI, *supra* note 48, at 11.

⁷⁷ *Id.* at 5–6.

⁷⁸ *Id.* at 11–12. *See also* *Authors Guild v. Google, Inc.*, 804 F.3d 202, 220 (2d Cir. 2015) (“[T]he copyright does not protect facts or ideas set forth in a work”).

⁷⁹ Gillotte, *supra* note 5, at 2684.

⁸⁰ Graphic Artists Guild, Comment Letter on Artificial Intelligence and Copyright 9–10 (Oct. 30, 2023), <https://perma.cc/2436-378S>.

⁸¹ *Id.* at 10.

prior to use in training, thus ensuring that creators receive compensation for use of their protected works.⁸²

Yet, many of the creative community's comments submitted in response to the Copyright Office's NOI seem to generally question AI developers' assertions that models do not copy or store copyrighted works.⁸³ However, if the AI developers' assertions are taken as true, neither comments submitted by the creative community nor existing legal scholarship appear to provide a concrete counterargument to the use of copyrighted works in training data as fair use. It would seem that, beyond the potential for licensing income, creators would be more concerned with infringing outputs rather than AI training.

III. CONCLUSION

The use of copyrighted materials in AI training should constitute fair use. Instead, the focus of copyright regulation should be on infringing outputs. If outputs are not infringing on copyrights, it is difficult to see why utilizing copyrighted works in AI training should be considered an infringing use. In fact, in suits against AI developers for copyright infringement, plaintiffs have argued that the injury incurred was that an output may infringe on their copyright.⁸⁴ These arguments have largely failed, as courts have found that an increased risk of future harm alone is insufficient to confer standing for damages.⁸⁵

If outputs infringe on copyrights, AI developers should be held vicariously liable where adequate safeguards were not employed. As previously discussed, AI developers have implemented a number of safeguards to minimize the risk of infringing outputs.⁸⁶ These safeguards include licensing agreements, opt-out programs, and output restrictions.⁸⁷ If effective, regulations requiring AI developers to implement such safeguards could be sufficient to minimize copyright concerns.

This scheme facilitates coherence between the federal government and AI developers while protecting the rights of copyright owners. Concerns over copyright infringement claims has driven data-rich companies to create their own AI models

⁸² Am. Soc'y of Composers, Authors and Publishers, Comment Letter on Artificial Intelligence and Copyright 4 (Oct. 31, 2023), <https://perma.cc/EP7T-YESD>.

⁸³ See, e.g., Graphic Artists Guild *supra* note 80 at 9–10.

⁸⁴ See, e.g., Doe 1 v. GitHub, Inc., No. 22-CV-06823-JST, 2023 WL 3449131 at 5 (N.D. Cal. May 11, 2023).

⁸⁵ *Id.*

⁸⁶ See *infra* Section II.B.1.

⁸⁷ *Id.*

trained off their IP.⁸⁸ In fact, several AI developers are facing copyright infringement claims for using unlicensed data to train AI models.⁸⁹ If successful, the AI innovation ecosystem may be transformed from the open-source driven process accessible with limited financial resources to a restricted system where reliable AI systems can only be developed by those with the most financial resources.⁹⁰ In other words, it may render the strategies of the federal government unachievable by impairing long-term AI innovation and the availability of safe and effective models.

Moreover, this scheme may benefit creators by encouraging AI developers to adopt additional safeguards. Rather than investing in licensing agreements that may disproportionately benefit larger creators, AI developers may invest in safeguarding the rights of all creators in its outputs. If copyright infringement of outputs are enforceable against developers, then developers utilizing copyrighted materials will likely be incentivized to develop adequate safeguards to minimize the risk of infringement.

Nevertheless, the protections necessary for creators expand far beyond this Article. Questions remain over a variety of issues such as who owns an AI output. Furthermore, the rise of “deep fake” AI outputs calls the sufficiency of the existing right of publicity doctrine into question.⁹¹ As the federal government has begun confronting these issues, the federal government and AI developers have been united in their commitment to driving AI innovation in a safe, accessible manner. If the Copyright Office determines that the use of copyrighted materials in AI training constitutes a copyright violation, this united approach to AI development will be significantly compromised. The benefits to creators through licensing income do not justify such an outcome, especially where alternative regulatory schemes can adequately address infringement concerns. Therefore, the Copyright Office should hold that only AI outputs can infringe on copyrights and hold AI developers vicariously liable if safeguards for minimizing infringement are not utilized.

⁸⁸ See, e.g., Brody Ford, *Adobe to Sell a New AI Subscription With Copyright Services (2)*, BLOOMBERG NEWS (June 8, 2023), <https://perma.cc/74SB-9TGW>.

⁸⁹ See, e.g., *Github*, 2023 WL 3449131; *Andersen v. Stability AI Ltd.*, No. 23-CV 00201, 2023 WL 7132064 (N.D. Cal. Oct. 30, 2023); *Getty Images (US) Inc. v. Stability AI, Inc.*, No. 23-CV-00135 (D. Del. filed Feb. 3, 2023); *Silverman v. OpenAI, Inc.*, No. 23-CV-03416 (N.D. Cal. filed July 7, 2023).

⁹⁰ See, e.g., Brody Ford, *Adobe to Sell a New AI Subscription With Copyright Services (2)*, BLOOMBERG LAW (June 8, 2023), <https://perma.cc/74SB-9TGW>.

⁹¹ Nate Lanxon, *Why are Deepfakes Everywhere? Can They Be Stopped?*, BLOOMBERG (Feb. 9, 2024), <https://perma.cc/3TJJ-MWJV>.