TECHNET
THE VOICE OF THE
INNOVATION ECONOMY

1420 New York Avenue NW, Suite 825
Washington, D.C. 20005
www.technet.org | @TechNetUpdate

October 30, 2023

U.S. Copyright Office
101 Independence Avenue SE
Washington, D.C. 20540

RE: Artificial Intelligence and Copyright

To Whom It May Concern:

TechNet is the national, bipartisan network of technology CEOs and senior executives that promotes the growth of the innovation economy by advocating a targeted policy agenda at the federal and 50-state level. TechNet's diverse membership includes dynamic American businesses ranging from startups to the most iconic companies on the planet and represents over 4.5 million employees and countless customers in the fields of information technology, artificial intelligence, e-commerce, the sharing and gig economies, advanced energy, transportation, cybersecurity, venture capital, and finance. As such, TechNet is uniquely positioned to advise the Copyright Office (the "Office") on the burgeoning field of artificial intelligence and its intersection with copyright law. Our comments reflect the majority position of TechNet's members but do not indicate unanimity.

Artificial intelligence may be the most transformative technological development since the creation of the Internet. Generative AI, in particular, has the potential to transform not only the creative industries, but other fields like software development, scientific research, healthcare, government administration, and education. The technology, at its core, promises to enable all people to create new content in any medium or any language, regardless of their skill level or ability. By breaking down barriers to creative expression, research and communication, Generative AI will unlock a more robust exchange of ideas and a more robust information economy. In this way, the promise of Generative AI aligns precisely with the goals of our intellectual property laws. This is a development copyright law should celebrate and encourage, not restrict. We urge policymakers to prioritize the preservation of core copyright law provisions that offer technology-neutral safeguards for legitimate rightsholders and innovators, which will enhance the competitiveness of American AI and creative industries on the global stage.

TechNet looks forward to working with the Office on an ongoing basis on these issues. To begin that process, TechNet is grateful for the opportunity to submit these responses to several of the questions raised in the Office's Notice of Inquiry. 88 Fed. Reg. 59942. Our responses to those questions are summarized below:

- First, we address "the use of copyrighted works to train AI models," *see* 88 Fed. Reg. at 59945, and explain that such use does not automatically violate copyright law. *See infra* Part B.

- Second, we address issues of upstream liability and the inefficiency of imposing liability on the developers of innovative tools that are capable of an extremely wide range of legitimate and socially beneficial uses. *See infra* Part C.

- Third, we endorse the Office's affirmation of the human authorship requirement as applied to AI-generated outputs, but nonetheless urge the Office to prioritize time-tested principles of authorship when addressing more difficult questions relating to works created through a combination of human ingenuity and AI tools. *See infra* Part D.

- Fourth, we address the possibility of a "compulsory licensing regime" for AI training, *see* 88 Fed. Reg. at 59947, and point out that any statutory licensing framework would run head-first into intractable economic and administrative problems. *See infra* Part E.

- Fifth, we respond to calls for copyright-adjacent legislation requiring AI developers to disclose their training data. We point out that such calls prove that AI training does not implicate any legitimate rightsholder interests, and that requiring the disclosure of trade-secret-protected information would undermine the ability of American companies to compete with our foreign adversaries. *See infra* Part F.

Training AI models does not inherently implicate rightsholders' copyright interests

"[C]opyright is not an inevitable, divine, or natural right that confers on authors the absolute ownership of their creations. It is designed rather to stimulate activity and progress in the arts for the intellectual enrichment of the public."[1] Copyright, in other words, is a public policy tool to assure authors that if they invest time and resources in creating new works of authorship, the law will, for limited times, grant to authors certain exclusive rights to those works. This is not a "special reward" for authors, but a "utilitarian" incentive mechanism to "encourage the production of works" for the benefit of the public.[2] In order to strike the appropriate balance between the rights of the copyright owner and those of the public, the Copyright Act not only grants exclusive rights to authors, but also imposes limitations and exceptions that limit the reach of those rights. For example, copyright has never required a follow-on creator to pay or get permission from an author before studying

[1] *Cambridge Univ. Press v. Patton*, 769 F.3d 1232, 1256 (11th Cir. 2014) (cleaned up); *see also Fox Film Corp. v. Doyal*, 286 U.S. 123, 127 (1932) ("The sole interest in the United States and the primary object in conferring the [copyright] monopoly lie in the general benefits derived by the public from the labors of authors.").
[2] *Patton*, 769 F.3d at 1238; *Google LLC v. Oracle Am., Inc.*, 141 S. Ct. 1183, 1195 (2021).

and learning from the ideas, facts, concepts, styles, and themes in an author's work.[3] In this way, copyright protects the ability of authors to market their works for profit, while also encouraging innovation, competition, and the free exchange of ideas. As the Supreme Court has recognized, "[t]he more artistic protection is favored, the more technological innovation may be discouraged; the administration of copyright law is an exercise in managing the tradeoff."

The goal of Generative AI is to help authors create *new* content—whether it consists of responses to user queries, longer pieces of text, visual works, music, or computer code. Encouraging the creation and dissemination of new expression, of course, is the very purpose of copyright law. No one could seriously contend that a technology whose sole function is to create new works is at cross-purposes with copyright law. The invention of the camera, for example, made it easier for people to create new visual works regardless of their skill with a pencil or paintbrush, opening the door to an entirely new and valuable medium of artistic expression. Similarly, Generative AI will make it easier for humans to create works in any medium, regardless of their skill level or ability,[4] thereby fostering a more robust and democratic creative economy. This is a development that copyright law should celebrate, not restrict. To say that Generative AI is a threat to the copyright framework is to fundamentally misunderstand both the intent of generative AI technologies—to augment, not replace, creators—and copyright's central purpose: not to provide a "special private benefit," but to encourage "the free flow of ideas" and "creative activity."[5]

Those who oppose AI on copyright grounds do not dispute these core principles. Instead, they raise two collateral attacks.[6] First, some have complained that Generative AI will create works that compete with their own. These artists worry that the low marginal cost of creating AI-generated works—combined with the impressive capabilities of these models—threatens their ability to compete in the

---

[3] *Eldred v. Ashcroft*, 537 U.S. 186, 219 (2003) ("[E]very idea, theory, and fact in a copyrighted work becomes instantly available for public exploitation at the moment of publication.").

[4] *See, e.g.*, Khari Johnson, *AI Could Change How Blind People See The World*, Wired (July 5, 2023), https://www.wired.com/story/ai-gpt4-could-change-how-blind-people-see-the-world/ (reporting on efforts to employ OpenAI's GPT-4 to assist the blind).

[5] *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417, 429 (1984).

[6] Some have also objected to Generative AI on the grounds that every output of an AI models is necessarily a "derivative work" because it is, in a colloquial sense, "based on" the works included in the training data used to create the model. *See, e.g.,* Complaint ¶ 59 *in Tremblay, et al. v. OpenAI, Inc., et al.*, No. 23-cv-03223, Dkt. 1 (N.D. Cal., filed June 28, 2023). But every Circuit to consider the question has reaffirmed the black-letter principle that "[i]n order to infringe the derivative right, there must be substantial similarity in protectable expression." 4 PATRY ON COPYRIGHT § 12:13 n.1 (collecting cases).

The same litigants have suggested that AI models are "themselves infringing derivative works" simply because they were trained on copyrighted content. *See, e.g.*, Complaint ¶ 56 *in Tremblay, et al. v. OpenAI, Inc., et al.*, No. 23-cv-03223, Dkt. 1 (N.D. Cal., filed June 28, 2023). But because those models do not "re-present [any] protected aspects of the original" works to users, they are not derivative works. *Authors Guild v. Google, Inc.*, 804 F.3d 202, 225–26 (2d Cir. 2015) (derivative work right does not apply to software program that "does not allow access in any substantial way to a [work]'s expressive content").

marketplace, and urge the Copyright Office or Congress to take some action to reduce that competition.

Copyright should not be used to restrict innovative technologies that may alter the competitive landscape by facilitating new modes of creativity. The use of copyright in this way would have condemned many valuable technological developments over the course of our history, including the invention of the camera (to which artists objected on the grounds that it would destroy art altogether[7]) and the advent of the modern recording industry (to which orchestra conductors objected on the grounds that it would destroy the demand for live performances.[8]) Those innovations have since enriched our creative economy many times over, both by opening the door to new media of expression and by generating massive new economic opportunities for creators. Generative AI is no different.[9] That the technology enables the creation of new works that might compete with the works of today's artists is not a reason to distort copyright into a tool of protectionism—particularly when doing so would have the collateral effect of impeding the use of Generative AI to solve some of humanity's "worst inequities"[10] and most difficult scientific and medical problems.[11] Current copyright law, in fact, does offer protection where there is true infringement, but to use copyright to throttle the creation of a technology for producing new works simply to protect existing copyright owners from competition would be to turn copyright policy on its head.[12]

The second attack on Generative AI is formalistic. It hinges entirely on the fact that, unlike the process of human learning, today's Generative AI models must make intermediate copies of content in order to learn from them. But the creation

---

[7] *See* Christine Height Farley, *The Lingering Effects of Copyright's Response to the Invention of Photography*, 65 U. Pitt. L. Rev. 385, 417–18 (2004).
[8] *See Revision of Copyright Laws, Hearings Before the House Comm. on Patents*, 74th Cong., 2d Sess. 680 (1936) (statement of Josef Pasternak).
[9] Artists across industries are turning to new AI tools to transform and optimize their art. *See e.g.*, Jazz Tangcay, *'Hollywood 2.0': How the Rise of AI Tools Like Runway Are Changing Filmmaking*, Variety (Feb. 22, 2023), https://variety.com/2023/artisans/news/artificial-intelligence-runway-everything-everywhere-all-at-once-1235532322/ (detailing how AI tools have optimized visual effects in filmmaking, such as in 'Everything Everywhere All at Once'); Bernard Marr, *The Amazing Ways Coca-Cola Uses Generative AI In Art And Advertising*, Forbes (Sept. 8, 2023), https://www.forbes.com/sites/bernardmarr/2023/09/08/the-amazing-ways-coca-cola-uses-generative-ai-in-art-and-advertising/?sh=5adb13da2874.
[10] Andrew Chow, *Why Bill Gates Believes Generative AI Will Be 'Revolutionary'*, Time (Mar. 21, 2023), https://time.com/6264801/bill-gates-ai/.
[11] Bernard Marr, *Breakthrough In Cancer Treatment: The Role of Generative AI in Drug Development*, Forbes (Sept. 20, 2023), https://www.forbes.com/sites/bernardmarr/2023/09/20/breakthrough-in-cancer-treatment-the-role-of-generative-ai-in-drug-development/?sh=555cfeb3b8c1 (reporting on efforts to use Generative AI tools in the "search for immunotherapy treatments").
[12] Indeed, several rightsholders who have vocally attacked Generative AI on copyright grounds have also publicly announced initiatives to build the technology into their own business plans—without any mention of the concern that the technology will unduly compete with traditional modes of artistic creation. *See, e.g.*, *YouTube Announces AI Music Principles and Launches YouTube Music AI Incubator with Artists, Songwriters and Producers from Universal Music Group*, Universal Music Group, (Aug. 21, 2023), https://www.universalmusic.com/youtube-announces-ai-music-principles-and-launches-youtube-music-ai-incubator-with-artists-songwriters-and-producers-from-universal-music-group/.

of intermediate copies in furtherance of the creation of a new and useful technological tool is not the kind of copying that violates copyright law. In fact, courts and commentators agree that this sort of intermediate, non-expressive copying constitutes a fair use.[13] Courts have recognized this principle across multiple contexts involving innovative technologies, including internet search engines, which "make[] and analyze[] a copy of each Web page [on the Internet]" in order to produce useful search results;[14] plagiarism-detection tools, which make unauthorized copies of term papers to spot cheating;[15] video game console emulators, where copies are necessary to enable interoperability;[16] and book search engines, where unauthorized copies of millions of books were made to facilitate new insights and keyword searching.[17] Other countries around the world also recognize that copyright law must leave room for legitimate and useful text and data mining.[18]

Ensuring that copyright does not expand beyond "its lawful bounds" to become a veto over new technologies is one of the core purposes of our fair use doctrine.[19] Indeed, as Judge Stephanos Bibas recently held, the intermediate copying of copyrighted content as a "step in the process of trying to develop a 'wholly new'" and non-infringing AI system is a "transformative" fair use, even if the resulting product "compet[es]" with the original.[20] The well-established body of "intermediate copying" case law on which that holding relied[21] has long been a bedrock part of our copyright framework that the Office itself has endorsed in its own Reports.[22]

AI companies have poured tens of billions of dollars of investment into Generative AI in reliance on that basic and fundamental point of law. That investment will yield huge advances in fields from physics to medicine to education, continuing economic growth, and tens or hundreds of millions of new jobs over the

---

[13] *See* Pamela Samuelson, *Unbundling Fair Uses*, 77 Fordham L. Rev. 2537, 2607 (2009) (citing *Sega Enterprises Ltd. v. Accolade, Inc.*, 977 F.2d 1510, 1514 (9th Cir. 1992)); *see also* Matthew Sag, *Copyright and Copy-Reliant Technology*, 103 NW. U. L. Rev. 1607, 1638 (2009); Mark A Lemley & Bryan Casey, *Fair Learning*, 99 Tex. L. Rev. 743, 785 (2020).
[14] *Field v. Google Inc.*, 412 F. Supp. 2d 1106, 1110–11 (D. Nev. 2006); *see also Perfect 10, Inc. v. Amazon.com, Inc.*, 508 F.3d 701, 723 (9th Cir. 2007).
[15] *See A.V. v. iParadigms, L.L.C.*, 562 F.3d 630, 634 (4th Cir. 2009).
[16] *See Sony Comp. Entertainment v. Connectix Corp.*, 203 F.3d 596, 599 (9th Cir. 2000); *see also Sega*, 977 F.2d at 1514.
[17] *See Authors Guild v. HathiTrust*, 755 F.3d 87, 103 (2d Cir. 2014); *Authors Guild v. Google, Inc.*, 804 F.3d 202, 207–08 (2d Cir. 2015).
[18] *See* Directive 2019/790, of the European Parliament and of the Council of 17 April 2019 on Copyright and Related Rights in the Digital Single Market and Amending Directives 96/9/EC and 2001/29/EC, 2019 O.J (L 130) 92, 113–14; *see also* Israeli Copyright Act § 19; Singaporean Copyright Act of 2021 §§ 19094; Japanese Copyright Act §§ 30-4, 47-4, and 47-5.
[19] *Google LLC v. Oracle Am., Inc.*, 141 S. Ct. 1183, 1198 (2021); *see also* H.R. Rep. 94-1476 at 65–66 (1976), *see also Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417, 429 (1984).
[20] *Thomson Reuters Enterprise Ctr. GmbH v. Ross Intelligence Inc.*, No. 20-cv-613, 2023 WL 6210901, at *8 (D. Del. Sept. 25, 2023).
[21] *See Sega Enterprises Ltd. v. Accolade, Inc.*, 977 F.2d 1510, 1527–28 (9th Cir. 1992); *Sony Computer Entm't, Inc. v. Connectix Corp.*, 203 F.3d 596, 608 (9th Cir. 2000).
[22] *See* U.S. Copyright Office, Report on Software-Enabled Consumer Products at 17–18 (2016), https://www.copyright.gov/policy/software/software-full-report.pdf.

next two years.[23] It will make countless people both more productive and more creative; it will certainly encourage, facilitate, and inspire the creation of more new, creative works than any technology since semiconductors and personal computers.[24] The desire by a limited set of advocates to upend copyright's foundational principles for short-term revenue generation would undermine law's goal of advancing creativity and technology in the long-term.

There is no basis to impose liability on AI developers for users' misdeeds

For the reasons discussed above, copying material for training purposes does not automatically violate copyright law. The question of whether the *outputs* of generative AI systems may violate copyright law is a separate and distinct question, and one that is likely to turn on the specific facts of each case. That analysis, however, must begin with a prior question: if an output does infringe copyright, *who* should be responsible?

Generative AI models, like many traditional software products, are general-purpose tools. Large language models, for example, can be used for a broad variety of tasks: to power analytical tools that analyze sentiment in a large set of online comments; to summarize or provide a critical review of a book; to translate speech from one language to another; to provide advice for a DIY project; to translate an idea for a short story into a piece of polished prose; or simply to engage in a conversation. The next generation of "multimodal" models—those that can understand and create outputs in any media, whether text, image, video, or sound—will have an even broader range of potential uses.[25] AI models, in other words, will be "Everything Machines." The scope of their potential uses will soon be as broad and infinite as the uses of a programming language or an operating system.

To impose liability on the creator of such a system for potentially infringing outputs makes no more sense than imposing liability on the creator of a word processor simply because someone used it to draft a work of fiction that infringes on a copyrighted book. In virtually all cases, it is the *user* of a Generative AI tool who supplies the prompt that dictates the resulting text or image. In the same way that the law might impose liability on an individual who uses Adobe Photoshop to create an unauthorized derivative of a copyrighted work (rather than imposing liability on

[23] Mohamed Kande & Murat Sonmez, *Don't Fear AI. It Will Lead to Long-Term Job Growth*, World Economic Forum (Oct. 26, 2020), https://www.weforum.org/agenda/2020/10/dont-fear-ai-it-will-lead-to-long-term-job-growth/ (arguing that "AI technology will create more jobs than it automates" and will lead to "an estimate $15.7 trillion, or 26% increase, in global GDP" by 2023).
[24] Alina Valyaeva, *AI Has Already Created As Many Images As Photographers Have Taken in 150 Years. Statistics for 2023*, EveryPixel (Aug. 15, 2023), https://journal.everypixel.com/ai-image-statistics.
[25] *See* Cade Metz, *What's the Future for A.I.?*, N.Y. Times (Mar. 31, 2023), https://www.nytimes.com/2023/03/31/technology/ai-chatbots-benefits-dangers.html ("A new wave of multimodal systems will juggle images, sounds and videos as well as text.").

Adobe for providing the tool), the law should impose liability on the *individuals* who use AI-powered image generators to achieve the same result.

Imposing liability on the tool-creator would also be inconsistent with well-established copyright doctrine, which protects the providers of neutral tools from copyright liability as long as those tools are "capable of substantial non-infringing uses."[26] Since the 1980s, that doctrine has protected "[i]nventors and entrepreneurs" from liability for the potential misdeeds of their uses, and has accordingly become one of the cornerstones of the pro-innovation environment to which we owe many of the technological developments that now power the American economy.[27]

This is not to say that developers of Generative AI tools should never be liable for infringing outputs—secondary liability doctrines may apply as they do to technology providers more generally. To be sure, because Generative AI is a novel technology unlike any other, and courts are best positioned in the first instance to develop new approaches for intermediating between "the respective values of supporting creative pursuits through copyright protection and promoting innovation in new [] technologies by limiting the incidence of liability for copyright infringement" in this new technological context.[28] But the Office should not endorse stringent liability rules that force AI developers to incur liability for every user action. Such rules will impose additional financial barriers to participating in this field, thereby discouraging the "[i]nventor[]" in her "garage" or "dorm room" from entering the industry in the first place.[29] They would also disincentivize developers from releasing foundational AI tools and models on an open-source basis, which, in turn, would make AI research far more costly for anyone other than highly-resourced laboratories and corporations.[30]

Moreover, most of today's AI developers have undertaken careful measures to prevent or discourage misuse of models or the creation of potentially infringing outputs. TechNet encourages developers to continue and implement these responsible development practices. These measures have included contractual use-restrictions, technical measures like prompt- and output-filters, and warnings that models are not intended to be used to infringe copyright or other intellectual property rights. These are salutary and important steps that developers should continue to explore as the industry develops. But the possibility that a user might circumvent or

---

[26] *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417, 442 (1984).
[27] *Metro-Goldwyn-Mayer Studios Inc. v. Grokster, Ltd.*, 545 U.S. 913, 956–60 (2005) (Breyer, J., concurring).
[28] *Id.* at 928 ("[T]he administration of copyright law is an exercise in managing the tradeoff" between these two values).
[29] *Id.* at 959 (Breyer, J., concurring).
[30] *See* FTC, *Generative AI Raises Competition Concerns*, Technology Blog (June 29, 2023), https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2023/06/generative-ai-raises-competition-concerns (noting the expense of "creating [] new model[s] from scratch" and the possibility that open-source releases of "base models" could "open up the playing field").

ignore those steps and use models to for an unlawful purpose should not trigger liability for the developers who created the model.

Copyright does not extend to works that are entirely AI-generated, but line-drawing will be difficult

The Office was correct in concluding that "copyright can protect only material that is the product of human creativity."[31] As the Office is aware, however, the line between a work that is the "product of human creativity" and a work that is purely "AI-generated" is sometimes difficult to draw. 88 Fed. Reg. 16191. The good news is that these difficult edge cases are likely to be rare in general and commercially significant in even fewer cases. In these rare scenarios, we believe that existing case law is sufficient for the Office's registration analysis. After all, most users of Generative AI will have no particular interest in the copyright status of most outputs, just as today nearly no one worries about the copyright status of their emails, instant messages, search results, and spreadsheets. And for those few creators who are interested in commercializing the outputs of Generative AI tools, there are numerous ways to combine their creative contributions with the contributions of an AI model in a manner that plainly yields a copyrightable final work. We encourage the Office to publicly state that human authorship utilizing Generative AI tools can be protectable under existing case law. For example, a human creator might use an AI model to edit, supplement, or improve a preliminary version of a work and then combine or arrange the AI-generated work with her own original creations. In these cases, the final work will support a copyright, giving the creator the ability to sell or license the resulting work and prevent infringing copying.

This is not to discount the difficulties that may arise in the rare cases where distinguishing the AI-generated elements from the human-generated elements will rise to the level of commercial importance. In resolving them, time-tested principles of authorship that our courts developed far before the advent of Generative AI can be brought to bear. As the Supreme Court explained 140 years ago in a seminal case concerning copyright's treatment of photography:

> The human intelligence . . . can produce nothing without material assistance; though man's help be a tool, a machine, another's hand, he does not the less produce a work of art, if he continues to exercise the faculties which are concerned in that art: sentiment, mind, taste. When the sculptor makes use of the precision compass, when the draughtsman employs the reducing mirror or the chambre claire, it is always the thought of the artist which directs the instrument, which guides and inspires the material means.[32]

---

[31] Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence, 88 Fed. Reg. 16190, 16191 (Mar. 16, 2023); *see also Thaler v. Perlmutter*, No. 22-cv-1564, 2023 WL 5333236, at *4–5 (D.D.C., Aug. 18, 2023) (same).

[32] *Sarony v. Burrow-Giles Lithographic Co.*, 17 Fed. Rep. 591, 599 (1883).

Statutory licensing schemes will not work in practice

During the recent hearing before the House Judiciary Subcommittee on Courts, Intellectual Property, and the Internet, Register Perlmutter noted several concerns with proposals for a statutory licensing regime to govern the use of copyrighted content to train AI models—including, most prominently, that it was unclear how such a regime "can be made feasible given the volume of works that would be involved."[33] This was entirely correct: while existing statutory licensing schemes might cover a universe of works in the tens of millions, a similar regime for works used for AI training purposes would need to administer royalty payments for many *billions* of works, representing many petabytes of data, most of which have no commercial value and were created without payment in mind. That would be orders of magnitude larger than any statutory or collective licensing framework ever deployed, covering a much more diverse set of works.

That scale creates an intractable economic problem. Any licensing framework that provided any significant compensation to individual authors would impose a massive and insurmountable barrier to AI development, as it requires tens of billions of individual works—and, accordingly, tens of billions of individual royalty payments—to train an effective model. This would bar anyone other than the largest technology companies from participating in the AI industry, effectively foreclosing the next wave of American start-ups. Even the largest technology companies would have a strong incentive to take their AI development efforts (particularly for AI model training)—and the tens of billions of dollars of investment capital underlying those efforts—to other jurisdictions with more innovation-friendly legal frameworks.

On the other hand, any statutory licensing scheme that imposed a less crippling financial obligation on the next generation of AI developers would mean that the resulting payments to individual authors would be miniscule. To an individual author, there would be no meaningful difference between the current copyright framework—under which they have no right to collect rents for the use of their works to train AI models, *see supra* Part B—and a licensing framework under which they receive royalty checks of a few cents each month. Such a scheme, with its attendant inefficiencies, neither benefits creators nor promotes the progress of science and the useful arts.

Worse still, any statutory licensing scheme would be impossible to administer. As Register Perlmutter recently explained, such a licensing framework for AI training would need to cover "all kinds of works"—including not only commercially available books, musical works, sound recordings, and professional photographs, but also other in-copyright works like internet comments, blog posts, online reviews, social media posts, amateur photographs, etc. The Office has already noted how "time-

[33] *See* Hearing on Oversight of the U.S. Copyright Office before House Judiciary Subcommittee on Courts, Intellectual Property, and the Internet at 1:11:00–:45, https://judiciary.house.gov/committee-activity/hearings/oversight-us-copyright-office (Sept. 27, 2023).

consuming, difficult, or even impossible" it is to locate the copyright owners for literary works.[34] However difficult it might be to locate the rightsholder for a book that has gone out print, that difficulty would pale in comparison to the challenge of identifying the rightful owners of the tens of billions of works used to train today's AI models.

In this connection, it is also important to emphasize that, where training is concerned, AI models generally derive their value from learning patterns and relationships that emerge from the relationships between billions, or even trillions, of examples. In other words, in stark contrast to most licensing contexts, market values based on expressive uses are no guide for non-expressive training uses. The corpus of millions of online comments posted on a large user-generated content site, for example, is likely to be more valuable for training LLM models than the combined archive of a major newspaper. Of course, these UGC platforms do not own the copyrights in the content of their users, and thus would not be a proper recipient of any collective licensing proceeds. Accordingly, in fairness, any collective or compulsory licensing scheme would have to allocate most of the value to the individual internet users, rather than large commercial publishers.

As a result, the amount of "unmatched" royalties under such a scheme would dwarf the amount of paid-out royalties by a significant margin. Most rightsholders would likely receive no remuneration whatsoever. Indeed, given the nature of AI training, which fundamentally is meant to extract facts and statistics from training data, developers can't determine how any particular work has been used, which makes any system of compensation that tries to value the contribution of each work fundamentally impossible. One of the most promising features of AI is that it finds relationships and patterns that we often cannot explain, much less find ourselves.

In summary, any legislation that would force AI developers to remunerate rightsholders for the use of their content to train AI models would ultimately fail to provide meaningful compensation to individual authors and, accordingly, fail to serve copyright's overall purpose of incentivizing the creation of new works. The only practical effect of a statutory licensing scheme would be to impose a substantial and burdensome tax on AI innovation—and, in turn, risk the transfer of American AI development overseas. This would in turn have severe detrimental impact on the long-term ability of creators and innovators alike to progress science and the useful arts as intended by law. U.S. Const. Art. 1, § 8, cl. 8. That kind of legislation has no rightful place in our copyright system.

Transparency obligations would unduly burden AI developers and handicap domestic AI development

In its Notice of Inquiry, the Office inquired about calls for copyright-adjacent legislation that would require AI developers to "disclose records regarding the

[34] U.S. Copyright Office, Report on Copyright and Digital Distance Education 41–43 (1999), https://www.copyright.gov/reports/de_rprt.pdf.

materials used to train their models" so that "copyright owners [could] determine whether their works have been used." 88 Fed. Reg. at 59947. That legislation would lead to a number of practical harms, as discussed below.

At a broader level, however, the fact that the works on which a model has been trained cannot be readily be determined by users of that model is proof that training works as designed—it is not meant to replicate the content of the works but to simply extract the unprotectable elements of those works. *See supra* Part B. We are aware of no other context in which a rightsholder has demanded remuneration for a use of a copyrighted work that is not only imperceptible, but impossible to detect.[35] This is precisely why disclosure requirements have never been a part of copyright law: because the only uses of a copyrighted work that implicate rightsholders' interests are those uses that actually communicate the expressive elements of a copyrighted work, in some form or another, to an audience, thereby usurping the rightsholder's ability to do or authorize the same.

Should such a broad requirement be adopted, there is no justification for imposing it on only one industry. Movie studios, for example, frequently use existing copyrighted works to inspire, guide, and develop projects. Should a filmmaker disclose every pre-existing work that was passed around in email, shared via cloud file storage lockers, or scanned or photocopied during a film's development? Should musicians be required to disclose every poem read on the web, every mp3 sent in an email, or every Instagram cover art inspiration that played a role in developing a hit song? Should software companies be required to disclose every scrap of code that was consulted during their development of a new software system? In every one of these cases, the works that may have been copied are not readily apparent to potentially aggrieved copyright owners, unless the expressive elements of those works are detectible in the final product.

Furthermore, legislation requiring broad disclosure of training data would discourage investment in critical AI technology and cripple American technology companies' ability to compete with their foreign counterparts. The creation of training datasets is a critical element of AI development.[36] Companies who build AI models devote substantial resources to selecting datasets that will yield effective, unbiased models. Which data is licensed, how it is obtained, and from whom, constitutes competitively valuable information. Still more resources are devoted to filtering data to (for example) remove duplicates and scrub potentially sensitive information, and to pre-processing data into a format suitable for model training. As a result, the "materials used to train [AI] models" are not merely a grab-bag of copyrighted works, but rather a highly curated set of data.

---

[35] *Cf. Authors Guild v. Google, Inc.*, 804 F.3d 202, 225 (2d Cir. 2015)(rejecting argument that the Google Books program is a derivative work, on the ground that the "program does not allow access in any substantial way to a book's expressive content" but "suppl[ies] information about that work").
[36] *See* Sang Michael Xie, et al., *Data Selection for Language Models via Importance Resampling* (Feb. 6, 2023), https://arxiv.org/pdf/2302.03169.pdf ("The choice of pretraining data is critical for the performance of language models.").

Forcing AI companies to disclose the contents of these datasets would, in effect, force the publication of valuable and otherwise confidential commercial secrets.[37] This may destroy the ability to claim trade secret protection for those assets under federal or state law.[38] In so doing, it would undermine the incentive to invest in new ways to compile, select, curate, and filter training data.[39] Worse still, it would require disclosure of these valuable trade secrets to foreign AI developers who may not be subject to similar disclosure requirements, thus handicapping United States technology companies and encouraging the next wave of American AI innovators to take their efforts overseas. Policymakers should instead focus on advancing the competitiveness of the United States AI and creative industries by protecting the core provisions of copyright law that offer technology-neutral protections to legitimate rightsholders and innovators.

Conclusion

Thank you for your attention to our views on this matter. We appreciate the opportunity to submit comments and provide feedback on the Office's study of copyright law and policy issues raised by AI. We stand ready to serve as a resource to you in your examination of this important issue.

Sincerely,

Carl Holshouser
Senior Vice President

[37] "Courts have long recognized that 'a trade secret can exist in a combination of characteristics and components, each of which, by itself, is in the public domain, but the unified process, design and operation of which, in unique combination, affords a competitive advantage and is a protectable secret.'" *AirFacts, Inc. v. de Amezaga*, 909 F.3d 84, 96 (4th Cir. 2018) (quoting *Imperial Chem. Indus. v. Nat'l Distillers & Chem. Corp.*, 342 F.2d 737, 742 (2d Cir. 1965)); *see also Mallet & Co. v. Lacayo*, 16 F.4th 364, 385 (3d Cir. 2021) ("A confidential compilation and organization of public information can amount to a trade secret.").

[38] *See, e.g.*, *KEMA, Inc. v. Koperwhats*, No. 09-cv-1587, 2010 WL 3464708, at *4 (N.D. Cal. Sept. 1, 2010) (disclosure of source code to administrative agency is "fatal to a trade secret claim").

[39] *See CVD, Inc. v. Raytheon Co.*, 769 F.2d 842, 850 (1st Cir. 1985) ("[T]he rationale behind [] trade secret law is to encourage invention, and to provide innovators with protection for the fruits of their labors.").