UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

IN RE: OPENAI, INC. COPYRIGHT INFRINGEMENT LITIGATION This Document Relates To: All Matters	25-md-3143 (SHS) (OTW) **STATEMENT OF INTEREST OF THE UNITED STATES**

The United States appears pursuant to 28 U.S.C. § 517[1] to inform the Court of its interest in this litigation.

I. THE INTEREST OF THE UNITED STATES

The United States has a strong interest in continuing to develop a robust and competitive artificial intelligence industry that sets the standard for the practice and procedure of AI use globally.[2] As the President stated in his Executive Order of January 23, 2025, "Removing Barriers to American Leadership in Artificial Intelligence," "[i]t is the policy of the United States to sustain and enhance America's global AI dominance in order to promote human flourishing, economic competitiveness, and national security."[3] As such, it is critical for the United States to "retain global leadership in artificial intelligence."[4] More recently, in his Executive Order of June 2, 2026, the President reiterated America's commitment to "encouraging AI innovation and accelerating responsible AI adoption across government and industry."[5] The United States thus has a strong interest in the question whether training AI models on written works constitutes "fair use" under copyright law.

---

[1] Section 517 provides, in pertinent part, that the Department of Justice may "attend to the interests of the United States in a suit pending in a court of the United States." The statute "contains no time limitation and does not require the Court's leave." *Gil v. Winn Dixie Stores, Inc.*, 242 F. Supp. 3d 1315, 1317 (S.D. Fla. 2017); *accord Karnoski v. Trump*, 2018 WL 4501484, at *2 (E.D. Mich. Sept. 20, 2018) (same); *United States ex rel. Lynch v. Univ. of Cincinnati Med. Ctr., LLC*, 2020 WL 1322790, at *4 (S.D. Ohio Mar. 20, 2020) (similar).

[2] The Government does not contend that any of the activities alleged or described in this litigation were authorized or consented to by the Government, or that any activities were for the benefit of the United States, as set forth by 28 U.S.C. § 1498.

[3] Executive Order, "Removing Barriers to American Leadership in Artificial Intelligence" (Jan. 23, 2025), https://perma.cc/QD6P-JEJE.

[4] *Id.*

[5] Executive Order, "Promoting Advanced Artificial Intelligence Innovation and Security" (June 2, 2026), https://perma.cc/TV79-9A55.

To start, effective integration of AI will be critical for U.S. national security capabilities. Several years ago, the Government Accountability Office (GAO) warned that "[f]ailure to adopt and effectively integrate AI technology could hinder national security" because AI can be used for tasks such as "analyzing intelligence information," "enhancing weapon systems (such as drones and robotic ships)," and "providing recommendations on the battlefield (such as where to target missile strikes)."[6] Over time, it has become even more apparent that "[a]dvanced AI capabilities make our Nation stronger, but also introduce new national security considerations" requiring that "the best and most secure technology is deployed to rapidly confront any and all threats to our country."[7] Of course, foreign adversaries have similar goals. Rules of law that make it significantly more difficult to develop a robust AI industry in the United States therefore threaten national security and give a competitive advantage to foreign adversaries who are not so encumbered.

The United States also has sovereign interests in the proper interpretation and application of federal law, including Article I, Section 8, Clause 8 of the United States Constitution (the Intellectual Property Clause) and the Copyright Act of 1976, 17 U.S.C. § 101 *et seq*. As the President's National Policy Framework for Artificial Intelligence explains, "American creators, publishers, and innovators should be protected from AI-generated outputs that infringe their protected content, without undermining lawful innovation and free expression."[8] In striking that balance, the President has encouraged Congress to consider enabling licensing frameworks or

---

[6] "How Artificial Intelligence Is Transforming National Security," GOV. ACCOUNTABILITY OFFICE (Apr. 19, 2022), https://perma.cc/62SS-H358.

[7] Executive Order, "Promoting Advanced Artificial Intelligence Innovation and Security," *supra* n.5.

[8] *See* Legislative Recommendations, National Policy Framework, Artificial Intelligence (March 2026), *available at* https://perma.cc/RJ8G-6FDC.

collective rights systems for rights holders (while suggesting that such legislation should not address when or whether such licensing is required), establishing a federal framework for protecting individuals from the unauthorized distribution or commercial use of AI-generated digital replicas of their identifiable attributes, and monitoring the development of copyright law to evaluate whether additional legislation is needed.[9] But consistent with the national interest in promoting innovation and free expression, the "training of AI models on copyrighted material," in and of itself, "does not violate copyright laws."[10]

In this litigation, the New York Times seeks to narrow fair-use doctrine to exclude the training of OpenAI's large language models (LLMs).[11] That result would be inconsistent with basic copyright law principles and severely hamper "the Progress of Science and useful Arts." U.S. CONST. art. I, § 8, cl. 8. As discussed below, publications—including independent and start-up media, in addition to the New York Times itself—are leveraging LLMs for journalistic and other purposes. And beyond the subject matter of this litigation, LLMs are already helping researchers across fields achieve major breakthroughs.[12] Constraining LLM development under a misunderstanding of fair use doctrine would thwart such creative and scientific progress while hindering American prosperity and economic mobility.

---

[9] *See id.*

[10] *Id.*

[11] To simplify, the United States refers only to Plaintiff the New York Times and Defendant OpenAI. The United States' legal arguments apply similarly to all parties in this litigation and the related cases, including book authors and publishers.

[12] *See, e.g.*, OpenAI, *An OpenAI model has disproved a central conjecture in discrete geometry*, https://openai.com/index/model-disproves-discrete-geometry-conjecture/ (last visited Aug. 31, 2026); Anthropic, *How Claude is accelerating protein design and analytical chemistry*, https://perma.cc/8HMS-EF5D (last visited Aug. 31, 2026).

It would cause other harms, too. An erroneous fair use ruling would hamper competition in the market for LLMs, because only the largest technology companies might have the capital necessary to pay licensing fees. And such licensing fees would disproportionately benefit legacy media outlets due to the sheer volume of their written publications. By contrast, if not hindered by a strained understanding of copyright law, LLMs can and should help level the playing field between mainstream and independent publishers, for several reasons. Authors with limited resources can use LLMs to compete (by, for example, using an LLM to generate an image to accompany an article—which otherwise might require a photographer or license). And LLMs can direct users to dissenting sources that offer contrary information or perspectives. It is not in the public's interest for the largest technology companies to have an oligopoly on LLM training due to licensing entry barriers that function primarily as large subsidies for old mainstream media companies.[13]

For all these reasons, the United States has a strong interest in this Court rejecting any argument that training LLMs on copyrighted texts violates copyright law.

[13] The United States takes no position on whether a licensing regime would be financially or logistically feasible given the breadth of works at issue, the liability that developers would face, and coordination difficulties. Whether or not requiring paid licensing would pose an existential threat to the American AI industry, it would at least hamper innovation and put American AI companies at a competitive disadvantage relative to competitors located in other countries that do not consistently respect U.S. intellectual property law. Notably, regardless of whether LLM model training on text articles constitutes fair use, both mainstream and independent publishers could enter (and have entered) into licensing agreements to provide developers with specialized access to real-time, pay-walled, proprietary, and other content and information.

II. STATEMENT

A. Relevant Background On AI LLM Training

As this Court has explained, Defendants, including OpenAI, develop LLMs—models that "can receive text prompts as inputs and generate natural language responses as outputs, which result from the LLM's prediction of the most likely string of text to follow the inputted string of text based on its training on billions of written works." *New York Times Co. v. Microsoft Corp.*, No. 1:23-cv-11195-SHS-OTW, ECF No. 514 at 6 (S.D.N.Y. April 4, 2025). LLMs are developed in several stages. *See id.* at 7-8. A developer "first collect[s] data, including plaintiffs' works," at the acquisition (or "collection" or "pre-training") stage. *Id.* at 8. This "involves collecting and storing a vast amount of content." *Id.* Next, a developer will "train their LLMs on that data through a process that feeds the data through the model." *Id.* Training involves making copies of the works comprising the training data, then converting that data into a set of numbers capable of representing many different meanings of words and their relationships to one another, which the LLM uses to learn linguistic patterns and make predictions. *See id.* At "the output stage," "LLMs respond to user queries" based on their training and any guardrails or filters the developer has set up. *See id.*

Each stage may present distinct questions of copyright law. The United States focuses on the question whether the use of copyrighted works at the training stage—by copying works in order to feed data into the model as learning material—constitutes fair use.

B. Legal Framework For Fair Use

The Intellectual Property Clause directs Congress "[t]o promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries." U.S. CONST. art. I, § 8, cl. 8. As that language makes clear, "[t]he primary objective of copyright is not to reward the labor of authors, but '[t]o promote the

Progress of Science and useful arts.'" *Feist Pubs., Inc. v. Rural Tel. Serv. Co., Inc.*, 499 U.S. 340, 349 (1991). Although "[t]he immediate effect of our copyright law is to secure a fair return for an author's creative labor," copyright law's "sole interest" and "primary object . . . lie in the general benefits derived by the public from the labor of authors." *Twentieth Century Music Corp. v. Aiken*, 422 U.S. 151, 156 (1975) (internal quotation marks omitted). "When technological change has rendered its literal terms ambiguous, the Copyright Act must be construed in light of this basic purpose." *Id.*

To achieve the constitutional objectives of the Intellectual Property Clause, the Copyright Act "strik[es] a balance between . . . rewarding authors' creations while also enabling others to build on that work." *Kirtsaeng v. John Wiley & Sons, Inc.*, 579 U.S. 197, 204 (2016). Although authors must have "a financial incentive to create informative, intellectually enriching works for public consumption," Courts have recognized "since shortly after the birth of copyright in England in 1710" that "giving authors *absolute* control over all copying from their works would tend in some circumstances to limit, rather than expand, public knowledge." *Authors Guild v. Google, Inc.*, 804 F.3d 202, 212 (2d Cir. 2015) (Leval, J.) (citing *Cary v. Kearsley*, 170 Eng. Rep. 679, 681 (1802) ("[O]ne must not put manacles upon science.")). Giving authors exclusive rights[14] encourages the creation and dissemination of expressive works; limiting those rights ensures that secondary users are permitted fair breathing room and latitude to facilitate further expression. *See Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 575 (1994) (recognizing the "inherent tension in the need simultaneously to protect copyrighted material and to allow others to build upon it").

[14] A valid copyright grants the owner "exclusive rights" to, among other things, "reproduce the copyrighted work in copies" and "distribute copies … of the copyrighted work to the public." 17 U.S.C. §§ 106(1), (3).

Uses of a copyrighted work are permitted when imposing infringement liability would "stifle the very creativity which [copyright] law is designed to foster." *Id*. at 577 (citation omitted).

A copyright owner's exclusive rights are thus subject to various exceptions and limitations. For example, "copyright assures authors the right to their original expression, but encourages others to build freely upon the ideas and information conveyed by a work." *Feist*, 499 U.S. at 349-50. "This principle, known as the idea/expression or fact/expression dichotomy, applies to all works of authorship." *Id.* at 350.

Relatedly, and as most relevant here, the "fair use" doctrine provides that certain secondary uses of a copyrighted work are "not an infringement." 17 U.S.C. § 107. Although fair use originated as "judge-made," Congress subsequently codified it. *Campbell*, 510 U.S. at 576. The statute continues a common-law tradition, which recognized that certain amounts and types of copying must occur to promote the purposes of the Intellectual Property Clause. *See id.* at 575 (citing U.S. CONST. art. I, § 8, cl.8). Fair use is an "equitable rule of reason that permits courts to avoid rigid application of the copyright statute when, on occasion, it would stifle the very creativity which that law is designed to foster." *Google LLC v. Oracle Am., Inc.*, 593 U.S. 1, 18 (2021).

The preamble of section 107 specifically recites six uses likely to result in a finding of fair use: "criticism, comment, news reporting, teaching . . ., scholarship, or research." But determining a "fair use" is "not to be simplified with bright-line rules, for the statute, like the doctrine it recognizes, calls for case-by-case analysis." *Campbell*, 510 U.S. at 576-77. As such, the statutory list is not exhaustive. *See* 17 U.S.C. § 107 (identifying purposes "such as" the listed set). The statute's legislative history confirms the same. *See Harper & Row Publishers v. Nation Enters*., 471 U.S. 539, 562 (1985); *Pac. & S. Co., Inc. v. Duncan*, 744 F.2d 1490, 1496 (11th Cir. 1984); *Cambridge Univ. Press v. Becker*, 863 F. Supp. 2d 1190, 1225 (N.D. Ga. 2012). The House Report

for section 107 indicates Congress's intent for a flexible inquiry that can adapt to new technologies and scenarios:

> [T]here is no disposition to freeze the doctrine in the statute, especially during a period of rapid technological change. Beyond a very broad statutory explanation of what fair use is and some of the criteria applicable to it, the courts must be free to adapt the doctrine to particular situations on a case-by-case basis.

H.R. Rep. No. 94–1476, 94th Cong., 2d Sess. 66 (1976); *see also id.* at 65 ("[S]ince the doctrine is an equitable rule of reason, no generally applicable definition is possible, and each case raising the question must be decided on its own facts.").

Congress and the courts have, however, established certain bedrock principles that guide the fair-use analysis. Section 107 identifies a nonexclusive list of factors "to be considered" in determining whether a particular use of a copyrighted work is "fair." 17 U.S.C. § 107. The enumerated factors are (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." *Id*. The fourth factor may also "take into account the public benefits the copying will likely produce." *Google*, 593 U.S. at 35.

As a general matter, the Section 107 factors cannot "be treated in isolation." *Campbell*, 510 U.S. at 578. Instead, "[a]ll are to be explored, and the results weighed together, in light of the purposes of copyright." *Id.*; *see also Google*, 593 U.S. at 19. That said, "the Supreme Court has made clear that some of the statute's four listed factors are more significant than others." *Authors Guild*, 804 F.3d at 213-14. At one point, the Supreme Court described the fourth factor as "undoubtedly the single most important." *Harper & Row*, 471 U.S. at 566. But more recently, "[i]n *Campbell*, the Court stressed also the importance of the first factor." *Authors Guild*, 804 F.3d

at 214 (citing *Campbell*, 510 U.S. at 591). "The more the appropriator is using the copied material for new, transformative purposes, the more it serves copyright's goal of enriching public knowledge and the less likely it is that the appropriation will serve as a substitute for the original or its plausible derivatives, shrinking the protected market opportunities of the copyrighted work." *Id.*

Importantly, because "[t]he same copying may be fair when used for one purpose but not another," "[t]he fair use provision . . . requires an analysis of the specific 'use' of a copyrighted work that is alleged to be 'an infringement.'" *Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith*, 598 U.S. 508, 533 (2023) (quoting 17 U.S.C. § 107). For example, in *Sony Corp. of America v. Universal City Studios, Inc.*, 464 U.S. 417 (1984), the Court classified recording television programs for private use as fair, but stressed that using the same recordings for other purposes—such as resale, public display, or commercial purposes—might produce a different result. *Id*. at 446, 451. Similarly, in *Campbell*, the Court explained that "[t]he use . . . of a copyrighted work to advertise a product, even in a parody, will be entitled to less indulgence under the first factor of the fair use enquiry than the sale of a parody for its own sake, let alone one performed a single time by students in school." 510 U.S. at 585. And in *Google*, the Court held that a precise copying of preexisting code was fair use given that Google's specific use of that code to create new products and platforms was "consistent with that creative progress that is the basic constitutional objective of copyright itself." 593 U.S. at 30.

The same use-by-use approach can apply within litigation. In *Authors Guild*, for example, Google had made digital copies of books, scanned the copies, and established a publicly available search function to determine whether a book contained a word or phrase. *See* 804 F.3d at 207. Users could also "see 'snippets' of text containing the searched-for terms." *Id.* And participating

libraries could download and retain digital copies of the books they had submitted to the program. *Id.* The Second Circuit carefully distinguished the search, snippet, and library download uses in its analysis (while approving all three as fair). *See id.* at 216-18, 221-223, 228-230.

C. Training Of LLMS On Written Works Is Exceedingly Transformative

The first statutory fair-use factor, the "purpose and character of the use," requires consideration of "whether the new work merely 'supersede[s] the objects' of the original creation ('supplanting' the original), or instead adds something new, with a further purpose or different character, altering the first with new expression, meaning, or message." *Campbell*, 510 U.S. at 578-79 (quoting *Folsom v. Marsh*, 9 F. Cas. 342, 348 (C.C.D. Mass. 1841) (No. 4,4901) (Story, J.) and *Harper & Row*, 471 U.S. at 562). The Supreme Court has described the latter type of use as "transformative." *Id.* "[T]ransformative uses tend to favor a fair use finding because a transformative use is one that communicates something new and different from the original or expands its utility, thus serving copyright's overall objective of contributing to public knowledge." *Authors Guild*, 804 F.3d at 214.

The copying of protected text articles as part of training an LLM is a use of a different kind or character that is "transformative—spectacularly so." *Bartz v. Anthropic PBC*, 787 F. Supp. 3d 1007, 1021 (N.D. Cal. 2025). The New York Times alleges that OpenAI's training results in a model that "predict[s] words that are likely to follow a given string of text based on the potentially billions of examples used to train" OpenAI's LLMs, such that the LLMs can subsequently generate original responses to a wide range of user inputs. *Microsoft Corp.*, No. 1:23-cv-11195-SHS-OTW, ECF 1677 ¶ 75 (Aug. 21, 2026). An OpenAI LLM thus uses the copyrighted work not to duplicate the work's expressive content, but as part of a process to learn and act on statistical patterns in written text, including vocabulary, syntax, and knowledge. The purpose of the copying (to build an intelligent, interactive model) differs in kind from the purpose of the copied work (to use

language to directly entertain or educate a reading audience). This use of text-based works to create an LLM engine for "innovative tools" that can "edit an email . . . , translate an excerpt from or into a foreign language, write a skit based on a hypothetical scenario, or do any number of other tasks" is undoubtedly "highly transformative." *Kadrey v. Meta Platforms, Inc.*, 788 F. Supp. 3d 1026, 1044 (N.D. Cal. 2025).

The Supreme Court's opinion in *Google* is instructive. The Court explained that Google (the secondary author) had written "new implementing code . . . designed to operate within" a new smartphone environment, and that although Google had copied preexisting code "precisely," it had done so for a purpose distinct from that of the original—"so that programmers who had learned an existing system could put their basic skills to use in a new one." 593 U.S. at 31. In the Court's view, this copying was fair because it facilitated programmers' new expression, and because a ban on copying would unduly impede Google's ability to exploit its own creative expression (the implementing code that it had written). *See id*. at 30–31. If that use was transformative, a developer's creation of a copy of a protected work as part of the process of teaching an LLM to recognize relationships between data and adapt to new information is as well.

To be sure, at the output (rather than training) stage, certain uses may not be transformative if the LLM reconstructs and disseminates an original copyrighted work. *See Authors Guild*, 804 F.3d at 222 (emphasizing that Google had "constructed the snippet feature in a manner that substantially protects against its serving as an effectively competing substitute" by displaying only a small "quantity of the copyrighted text"). But because of the required use-by-use analysis, the development of a sophisticated tool that can understand patterns and make predictions in response to user inputs must be distinguished from how persons might hypothetically or anomalously use that tool to generate certain outputs. *Bartz*, 787 F. Supp. 3d at 1021 (delineating arguments

"challeng[ing] only the inputs, not the outputs, of these LLMs"). The New York Times alleges that OpenAI's LLMs can be used to generate portions of its works. *Microsoft Corp.*, No. 1:23-cv-11195-SHS-OTW, ECF 1677 ¶¶ 80, 99, 100 (Aug. 21, 2026). But whatever legal questions certain output uses might raise, that should not bear on the transformative nature (or any other aspect) of the training use.[15]

In sum, the use of copies to train LLMs is extraordinarily transformative. That OpenAI uses its LLMs for commercial purposes does not move the needle. "Nearly all of the illustrative uses listed in the preamble paragraph of § 107 . . . are generally conducted for profit." *Campbell*, 510 U.S. at 584. And "the more transformative the [secondary] work, the less will be the significance of other factors, like commercialism." *Authors Guild*, 804 F.3d at 219 (quoting *Campbell*, 510 U.S. at 579).[16]

---

[15] In all events, any prospective remedy for infringement would need to be limited to what is necessary to afford plaintiffs complete relief. *See, e.g.*, *Trump v. CASA, Inc.*, 606 U.S. 831, 852-54 (2025). A tiny sliver of anomalous reconstructive outputs would not support a remedy that cuts off or threatens massive liability for LLM output uses generally, let alone distinct LLM training uses. OpenAI states that it has taken steps to prevent the substantial reproduction of the protected aspects of the written works on which it trains. *See, e.g.*, No. 1:25-md-03143-SHS-OTW, ECF No. 276 slide 48 (June 27, 2025).

[16] As is often the case, the second and third fair-use factors are unlikely to play a major role in the analysis. But the second factor (the "nature of the copyrighted work") likely supports a fair-use ruling because the training use's "transformative purpose . . . inevitably involves the second factor as well." *Authors Guild*, 804 F.3d at 220 (internal quotation marks omitted). The third factor ("the amount and substantiality of the portion used in relation to the copyrighted work as a whole") likely supports a fair-use ruling, too. Although LLM training involves the copying of entire works at an early step, "[w]hat matters . . . is not so much the amount and substantiality of the portion used *in making a copy*, but rather the amount and substantiality of *what is thereby made accessible* to a public for which it may serve as a competing substitute." *Id.* at 221-22. Training an LLM, in and of itself, does not make any copied works accessible to the public. Even at the distinct output stage, at most a very small fraction of outputs would make protected aspects of those works accessible. As noted, particular outputs can be assessed case-by-case.

D. Training Of AI Models Does Not Hamper The Potential Market For Or Value Of The Copyrighted Work In The Relevant Sense

The fourth fair use factor, "the effect of the use upon the potential market for or value of the copyrighted work," 17 U.S.C. § 107, supports the conclusion that OpenAI's model training using New York Times articles is fair use. The fact that OpenAI's training use is (more than) "highly transformative . . . favors a finding of fair use . . . unless the value of its transformative purpose is overcome by its providing text in a manner that offers a competing substitute" for New York Times works. *Authors Guild*, 804 F.3d at 218. A copy used as part of the training process does not "serve as a substitute for the original" or "shrink[] the protected market opportunities of the copyrighted work" in the relevant sense—let alone in a manner that would overcome the value of the use's extraordinarily transformative purpose. *Id.* at 214.

To "affect the market for the original in a way cognizable under this factor," a use must "act[] as a substitute for" the copyrighted work. *Campbell*, 510 U.S. at 591; *see also Harper & Row*, 471 U.S. at 567 (the use must "materially impair the marketability of the work which is copied"). The question is not whether the use could cause "some loss of sales." *Authors Guild*, 804 F.3d at 224. What matters is "significant *substitutive* competition," not competition generally. *Id.* at 219 (emphasis added). Because copyright "protects only the author's manner of expression," the question is whether the use "could provide a significant substitute for the purchase of the author's [work]." *Id.* at 224-25. That hinges on whether the copying is "done in a manner that results in widespread revelation of sufficiently significant portions of the original as to make available a significantly competing substitute." *Id.* at 223. After all, to succeed on any "claim of infringement under the Copyright Act, . . . [a] plaintiff must usually show not only a similarity but a 'substantial' similarity between the allegedly infringing work and the original elements of his own copyrighted work." *Goldsmith*, 598 U.S. at 557 (Gorsuch, J., concurring) (citing 4 Nimmer

on Copyright § 13.03[A] (2023)); *accord Yurman Design, Inc. v. PAJ, Inc.*, 262 F.3d 101, 110 (2d Cir. 2001) (similar).

Using a copyrighted work to train an LLM—without more—generally does not result in this sort of substitution because it does not "reveal[]" a significant amount of original "authorial expression." *Authors Guild*, 804 F.3d at 224; *see also, e.g.*, *Bartz*, 787 F. Supp. 3d at 1031 ("[T]raining LLMs did not result in any exact copies nor even infringing knockoffs of their works being provided to the public."). In fact, training does not reveal anything to the public at all—it simply creates a copy of a protected work in order to teach an LLM to recognize relationships between data and adapt to new information. The potential for future outputs that might cause market harm is simply not relevant to evaluating an LLM training use under the required use-by-use analysis.

To be clear, even LLM outputs that compete with text articles—without reproducing or substantially resembling protected aspects of text articles—would not be substantially similar to, or substitutes for, copyrighted works in the relevant sense. When outputs "copy no protected elements of the original work, much less significant portions," they cannot cause the relevant form of market harm just because they happen to be "in the same genre or category of works" as the original, given that "a genre is an uncopyrightable idea or method of expression." Edward Lee, *Copyright Dilution Under Constitutional Scrutiny*, 25 Chi.-Kent J. Intell. Prop. 1, 6 (2026) (citing *Peters v. West*, 692 F.3d 629, 636 (7th Cir. 2012) ("[N]o poet can claim copyright protection in the form of a sonnet or a limerick.")); *accord Abdin v. CBS Broad. Inc.*, 971 F.3d 57, 70 (2d Cir. 2020) (no infringement where "an independent comparison of the works reveals that there is no substantial similarity between the protectible features of [the original]" and the secondary use). Whether a particular output or category of outputs is substantially similar to the copyrighted work,

and whether any substantially similar reproduction might be a significantly competing substitute, are distinct questions involving distinct "challenged use[s]" (and potentially additional legal questions). *Harper & Row*, 471 U.S. at 568. But outputs lacking in substantial similarity cannot cause the sort of market harm that is cognizable in the fair-use analysis.

Contrary dicta in a Northern District of California decision misapplies copyright principles to LLM training. *See Kadrey*, 788 F. Supp. 3d at 1052-57. Without the benefit of briefing, the *Kadrey* court adopted an "indirect substitution" theory of "market dilution" based on the potential "use [of] LLMs to create books and then sell them, competing with books written by human authors." *Id.* at 1052, 1054 (internal quotation marks omitted); *see also id.* at 1036 (noting "the plaintiffs barely g[a]ve this issue lip service"). The court acknowledged that "in a more typical case, an original work is being compared to a single secondary work," and whether the works are "similar" is relevant. *Id.* at 1054. The court also appeared to acknowledge that its theory of indirect market dilution has "never made a difference in a case before." *Id.* at 1055. But in the court's view, similarity between an original work and a secondary use is irrelevant in this context, because LLMs "are trained on the creative expression in copyrighted books" as an initial matter, and LLM outputs have the "potential to flood the market with competing works." *Id.* at 1054-55. The court thus stated that LLM developers "will generally need to pay copyright holders for the right to use their materials" even for training, because LLMs will often have "the ability to severely harm the market for the works being copied, and thus severely undermine the incentive for human beings to create." *Id.* at 1035-1036.

Even putting aside the fact that human beings *use LLMs to create*, the *Kadrey* court's application of the fourth fair-use factor is deeply flawed. Most obviously, the court improperly collapsed LLM training and LLM outputs into a single continuous use, then applied a capacious,

genre-level understanding of "substitution." But LLM training is its own use. It does not involve any substitution. And even when it comes to LLM outputs competing with text articles, "the type of loss of sale envisioned . . . will generally occur in relation to interests that are not protected by the copyright." *See Authors Guild*, 804 F.3d at 224. In other words, the *Kadrey* court's broad market dilution theory overlooks arguably the most basic element of copyright infringement—whether the "works are substantially similar," *Goldsmith*, 598 U.S. at 525—by conflating training (which requires copying of entire works, but no public access) with outputs (which the public may access, but which will often if not always lack substantial similarity).[17]

This type of logic would have problematic implications for copyright law generally. To illustrate, when she was a teenager, Joan Didion "would type out" Ernest Hemingway's "stories to learn how the sentences worked," and as a result she considered him the greatest influence on her writing. *See* Linda Kuehl, *Joan Didion, the Art of Fiction No. 71*, The Paris Review (Issue 74, Fall-Winter 1978).[18] By the *Kadrey* court's logic, Didion should have incurred liability to Hemingway every time she published a piece, because the process by which she trained herself and the process by which she produced works was all one use, and her works competed with those

---

[17] The *Kadrey* court did not cite any case supporting such a theory. But the Register of Copyrights, who is currently challenging her removal, appeared to endorse a similar theory in a report. *See* U.S. COPYRIGHT OFFICE, COPYRIGHT AND ARTIFICIAL INTELLIGENCE PART 3: GENERATIVE AI TRAINING 64 (May 2025), https://perma.cc/Q2TQ-Y37B. Her understanding does not warrant deference. *Loper Bright Enters. v. Raimondo*, 603 U.S. 369, 412 (2024). And the Register's threadbare reasoning ignored all the caselaw emphasizing the required use-by-use analysis and discussing what sort of "effect" upon the potential market is cognizable. *Id.* Although the question "whether unrestricted and widespread conduct of the sort engaged in by the defendant would result in substantially adverse impact on the potential market for the original" can be relevant, U.S. COPYRIGHT OFFICE, COPYRIGHT AND ARTIFICIAL INTELLIGENCE PART 3: GENERATIVE AI TRAINING 65 n.373 (quoting *Campbell*, 510 U.S. at 590), such competition must both arise from the specific use in question and involve the requisite substitutive similarity.

[18] *Available at* https://www.theparisreview.org/interviews/3439/the-art-of-fiction-no-71-joan-didion.

of other authors in the market for literature. But "to make anyone pay specifically for the use of a book . . . each time they later draw upon it when writing new things in new ways would be unthinkable." *Bartz*, 787 F. Supp. 3d at 1021.

Of course, fair use is "a flexible doctrine that takes account of 'significant changes in technology.'" *Kadrey*, 788 F. Supp. 3d at 1055 (quoting *Google*, 593 U.S. at 19). But technological change cannot justify "turn[ing] . . . fair use analysis on its head." Lee, *supra*, at 6. When the Supreme Court must "apply traditional copyright concepts in th[e] technological world," it does "not change[] the nature of those concepts," but rather "applie[s] them to this different kind of" circumstance. *Google*, 593 U.S. at 40. If anything, contrary to the *Kadrey* court, the radically transformative nature of the technology at issue supports a fair-use finding. Human beings create original works using LLMs. And copyright law is not just (or even primarily) about protecting financial incentives for original authorship anyway. Instead, "the ultimate, primary intended beneficiary is the public." *Authors Guild*, 804 F.3d at 212. A market dilution analysis that conflates uses and focuses on potential financial harm due to generalized competition is untethered from that purpose. Whether this new technology warrants a wholesale rewriting of copyright principles is a policy question best left to the People and their elected representatives, subject to any applicable constitutional constraints.

Additionally, focusing on general competitive harm that could arise from new technology ignores that the fourth factor "take[s] into account the public benefits the copying will likely produce," whether such benefits are "related to copyright's concern for . . . creative expression," and whether they are "comparatively important, or unimportant, when compared with dollar amounts likely lost (taking into account as well the nature of the source of the loss)." *Google*, 593 U.S. at 35-36. For example, where "the reimplementation of a user interface allow[ed] creative

new computer code to more easily enter the market," the Supreme Court held that "the risk of creativity-related harms to the public" that would result from a contrary ruling meant that "this fourth factor—market effects—also weighs in favor of fair use." *Id.* at 40. When a secondary use can lead to "creative improvements, new applications, and new uses," treating copyright law as "a lock limiting . . . future creativity . . . would interfere with, not further, copyright's basic creativity objectives." *Id.* at 39. That reasoning applies with even more force here. A ruling that renders LLM training subject to copyright infringement liability would threaten to "stifle the very creativity which that law is designed to foster." *Campbell*, 510 U.S. at 577.

At a minimum, the creative possibilities and public benefits that LLM training advances far outweigh any competitive harm (even assuming such harm is cognizable). The entire reason that AI models are rapidly reshaping the economy and national security is that they help people, including those working in creative fields, make things and get things done. LLMs can inspire or help someone to write a story, compose a song, write a movie script, or produce any other kind of art. Indeed, authors at the New York Times itself are using LLMs to help them "conceptualize and edit" articles.[19] Independent and start-up publications, as well as ordinary people, can too. For example, an independent writer used an LLM and his background as a physics teacher to offer a contrarian perspective about data center water usage and critique the New York Times.[20] Such

---

[19] Maggie Harrison Dupre, *New York Times Issues Stern Warning to its Freelance Writers about AI Use*, FUTURISM (May 12, 2026), *available at* https://perma.cc/KY58-D2HB.

[20] *See* Andy Masley, *Using ChatGPT is not bad for the environment* (Jan. 13, 2025), *available at* https://perma.cc/BG4W-YPVM; *see also* Blake Dodge and Harris Sockel, *The Data Center Water Crisis Isn't Real*, Pirate Wires (Dec. 18, 2025), *available at* https://www.piratewires.com/p/andy-masley-ai-water-crisis-isnt-real (relying on Masley's research and commentary while discussing a July 2025 New York Times article titled "Their Water Taps Ran Dry When Meta Built Next Door," which purported to show that "data centers . . . guzzle up water" by pointing to a couple who experienced residential water problems before a nearby data center was operational).

creative benefits often overlap with national security interests, because LLMs can, for example, help national security officials analyze and draw inferences from (unprotected) real-world facts such as those conveyed in New York Times articles. *See Feist*, 499 U.S. at 349.

Given all these considerations, the fourth factor heavily favors fair use. At a minimum, any effect on the market for the copyrighted works cannot overcome the significant, world-changing value of the training use's transformative purpose. *See Authors Guild*, 804 F.3d at 218.

*

The fair-use inquiry hinges on the specific facts and uses at issue in each case. But it would be problematic—and legally incorrect—to impose broad copyright liability that would generally render training of AI models impermissible without licensing. LLM training is "consistent with that creative 'progress' that is the basic constitutional objective of copyright itself." *Google*, 593 U.S. at 30.

Respectfully submitted,

STANLEY E. WOODWARD, JR.
Associate Attorney General

BRETT SHUMATE
Assistant Attorney General, Civil Division

By: /s/ Michael Weisbuch
MICHAEL WEISBUCH
Senior Counsel to the Associate Attorney General
United States Department of Justice
950 Pennsylvania Ave., NW
Washington, DC 20530
(202) 322-2450
Michael.Weisbuch@usdoj.gov

Attorneys for the United States of America

September 1, 2026