AI PROGRESS

AI MODELS: ADDRESSING MISCONCEPTIONS ABOUT TRAINING AND COPYRIGHT

BY ANNA CHAUVET AND
KARTHIK KUMAR, PH.D.

AI PROGRESS

WHY U.S. COPYRIGHT LAW SUPPORTS AI INNOVATION

EXECUTIVE SUMMARY

As debates continue around copyright and generative AI, it's critical to separate fact from fiction about how these AI models are trained and what the law protects. Much of the current conversation is fueled by misconceptions about how AI models use training data. And the stakes couldn't be higher: Restricting AI model development due to misinterpretations of U.S. copyright law would hinder innovation, stall progress, and undercut America's global leadership in AI.

Generative AI — especially large language models (LLMs) — are trained on large datasets to learn how to generate accurate, relevant responses. Unlike databases, these models do not need to store or systematically retrieve content. Instead, they generate new material based on patterns, structures, and word relationships learned during training. Some critics argue this process violates U.S. copyright law. But the law protects specific creative expression, not underlying facts or patterns — which are the elements used by AI models.

U.S. copyright law — particularly the fair use doctrine — provides clear protections for this kind of transformative use. Courts have consistently recognized that innovative technologies qualify as fair use when they use copyrighted materials in new ways that don't substitute for the original works in the marketplace. Recently, two courts have found that generative AI training meets this standard.

KEY TAKEAWAYS

AI MODELS GENERATE NEW WORKS BASED ON LEARNED PATTERNS — NOT STORED DATA.

They do not need to retrieve or reproduce original training material, but instead rely on learned statistical relationships or patterns from across a broad body of content to generate responses.

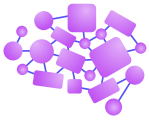
COPYRIGHT LAW PROTECTS EXPRESSION, NOT IDEAS.

U.S. copyright law does not protect ideas, facts, or patterns revealed in material, just the specific way those things are conveyed. Similarly, the law doesn't protect against new competition — only against copying existing expression.

COURTS BACK USING COPYRIGHTED MATERIALS IN GENERATIVE AI TRAINING.

Courts have found that transformative uses — including generative AI training on whole books or datasets — are lawful when they serve a new purpose and don't compete with the original works.

EFFORTS TO RESTRICT AI TRAINING THROUGH COPYRIGHT LAW THREATEN INNOVATION.



Limiting access to training data would stall the development of AI technologies and run counter to the goals of copyright, which seeks to promote creativity and progress.

AI PROGRESS

ABOUT THE AUTHORS

ANNA CHAUVET

Anna Chauvet is a partner at Finnegan and leader of the firm's copyright practice. She brings deep experience in copyright litigation, licensing, and policy, with a focus on emerging issues at the intersection of AI and copyright. A former associate general counsel at the U.S. Copyright Office, Anna regularly advises on legislative and regulatory matters and is active in shaping the field through leadership roles with the Intellectual Property Owners Association's Copyright Committee and Licensing Executives Society International's Copyright Committee. She is a frequent speaker before industry, government, and academic audiences.

KARTHIK KUMAR, PH.D.

Karthik Kumar, Ph.D. is a partner at Finnegan specializing in competitor IP litigation and strategic counseling on emerging technologies. Recognized by IAM Strategy 300 as a global IP leader, he represents clients before federal courts, the USPTO, and the ITC, and advises leading tech companies on global patent portfolios spanning AI/ML, AR/VR, quantum computing, 5G/6G, and beyond. A co-lead of Finnegan's AI/ML and AR/VR working groups, Karthik brings 15 years of cross-sector experience and is also a co-inventor on nine U.S. patents with a background in plasma physics, quantum electronics, and optics.

AI Models: Addressing Misconceptions About Training and Copyright

By Anna Chauvet and Karthik Kumar, Ph.D.

Artificial intelligence (AI) is a foundational and transformative technology with the power to drive scientific discoveries, generate new economic opportunities for individual businesses, and transform industries.[1] Due to its potential to fuel macroeconomic expansion and strengthen national security,[2] the United States government has made fostering AI innovation a strategic priority of the highest order. Such advancement in AI fundamentally depends on the ability of the models to be trained on large quantities of data. The technical process of "learning" for AI models means being able to derive patterns, structures, and relationships from across a broad body of data so that the models can predict and generate original content in response to user queries. The larger the scope and breadth of data used to train an AI model, the more accurate and robust its responses will be.

Misunderstandings about AI model training may lead to attempts to limit or impede the use of data for that purpose, thereby risking the United States' technological progress in AI. For example, some groups are seeking to ban AI training on copyrighted material without a license, despite AI training falling squarely within the bounds of copyright law's fair use doctrine (as discussed below). Others have sought new statutory requirements compelling disclosure of all data on which a model is trained, notwithstanding such information being closely guarded trade secrets. If adopted, these proposals would create unnecessary barriers to progress, hinder economic growth, and weaken America's competitive edge in the global AI landscape.

Accordingly, in this paper, we address certain mischaracterizations about AI model training processes and content generation. We also discuss how United States copyright law, particularly the fair use doctrine, provides the necessary flexibility to address many scenarios likely to arise with AI. Courts have decades of experience in applying the fair use doctrine to new technologies, and AI is no different. Finally, we discuss how, consistent with legal precedent, training AI models on copyrighted works constitutes a transformative and fair use, as it furthers the development of new and innovative technologies and encourages the creation of new expressive content. While this paper, for illustrative purposes, will focus on large language models (LLMs) — a type of AI system trained to understand and generate human language — our conclusions apply to AI development more generally.

---

[1] When discussing AI in this paper, we are referring to generative AI models (i.e., AI models with the ability to create new content, such as text, images, audio, and video, based on users' prompts or inputs).

[2] Goldman Sachs projects that AI will have a measurable effect on United States GDP by 2027 and may contribute to a 7% annual increase in global GDP over the next decade. *AI may start to boost US GDP in 2027*, GOLDMAN SACHS (Nov. 7, 2023), https://www.goldmansachs.com/insights/articles/ai-may-start-to-boost-us-gdp-in-2027.html. Similarly, Bank of America predicts that AI will drive employment growth in fields such as aerospace, information technology, education, and health care. *Artificial intelligence: A* real *game changer*, BANK OF AM., https://www.privatebank.bankofamerica.com/articles/economic-impact-of-ai.html (last visited July 2, 2025).

I. AI Models Generate New Content Based on Statistical Relationships Learned During Training

AI models differ fundamentally from databases because they are not designed to store or systematically retrieve content. Rather than scripting responses from stored content in response to user prompts, AI models spontaneously generate outputs.

Specifically, AI models generate responses based on statistical relationships that they learned from training and an optimization process that determines the most contextually appropriate response. For example, an AI developer building an LLM trains the model on a vast dataset of text from which the model can derive statistical information reflecting the relationships between different words and the grammatical rules governing speech patterns and language. The technical process of "learning" for LLMs involves deriving language patterns, structures, and relationships from across a broad body of training data, which enables them to generate responses using probabilistic modeling — that is, by predicting the most likely next word based on the statistical patterns learned during training.

An LLM's architecture consists of multiple layers of artificial neurons (mathematical functions that simulate the way the human brain processes information), each with adjustable parameters (to help determine how the model interprets and generates text) that are fine-tuned during training so the model can better recognize word patterns and relationships, and thus predict and generate content based on learned patterns.[3] The training process begins with the model processing millions, if not billions, of text sequences.[4] This involves analyzing vast amounts of raw text from training data (from which low quality, redundant, or harmful content was previously filtered out), and breaking it down into "tokens."[5] "Tokens" are small units — like words, sub-words, or even characters — that the model can process.[6] Through training, each token is assigned a numerical representation ("embedding") that captures its meaning within a structured vocabulary.[7] Embeddings are a way of transforming complex, human-readable data (like words) into numbers that a computer can process. In LLMs, words are mapped in a multi-dimensional space, and their place within that space is defined with a set of numbers — called a "vector" — with each number reflecting the distance along one of the axes that define that space. The vector reflects characteristics about the token, such as its semantic meaning or syntactic role.

---

[3] A. Feder Cooper et al., *Report of the 1st Workshop on Generative AI and Law,* ARXIV (Dec. 3, 2023), https://arxiv.org/pdf/2311.06477.

[4] Zhiqiang Shen et al., *Understanding Data Combinations for LLM Training*, ARXIV (May 9, 2024), https://arxiv.org/pdf/2309.10818.

[5] A. Feder Cooper et al., *Report of the 1st Workshop on Generative AI and Law,* ARXIV (Dec. 3, 2023), https://arxiv.org/pdf/2311.06477.

[6] *Id.*

[7] *Id.*

For example, if the sentence "The princesses had an unbelievable victory in yesterday's game" appeared in training data, the LLM would break it into tokens like "the," "princesses," and "victory." The LLM would then reduce words to their base or root form, to help the model treat different forms of a word as the same concept (e.g., "unbelievable" → "un" and "believable"; "princesses" → "princess"; "yesterday's" → "yesterday"; "un" → "not"; and "believable" → "belief"). These tokens form the basis for embeddings — high-dimensional vectors that encode relationships between words and represent what the model learns during training.[8] For example, when encountering the word "princess" in training data, the model assigns it a numerical value based on contextual usage, which is then positioned within a mathematical map of language.[9] In this space, "princess" might be positioned close to "queen" or "king," while words like "chair" and "bed" would be farther away (*see* Figure 1).

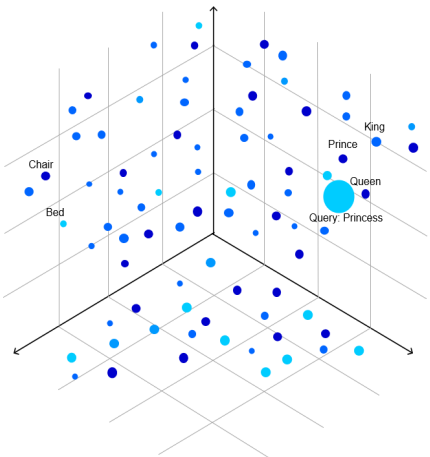


Figure 1

In another example, suppose the sentence "Emily watched the bat fly at night" appeared in the training data. In processing that sentence, the model would adjust the numbers in the vectors for the words "Emily," "watch," "the," "bat," "fly," "at," and "night" to reflect that those words appeared together in a sentence. After encountering thousands of other sentences or sentence fragments containing some combination of the words "bat," "fly," and "night" — and far fewer sentences combining those words with "Emily" or "watch" — the model's vectors will begin to reflect the semantic proximity (or distance) of those words.

---

[8] Ayesha Sidhikha, *Evolution of Language Representation Techniques: A Journey from BoW to GPT*, MEDIUM (Nov. 4, 2024), https://medium.com/@ayeshasidhikha188/evolution-of-language-representation-techniques-a-journey-from-bow-to-gpt-d3086c0168d1.

[9] *Integrating Vector Databases with LLM: Techniques & Challenges*, AIRBYTE (Aug. 29, 2024), https://airbyte.com/data-engineering-resources/integrating-vector-databases-with-llm.

Similarly, after the model encounters many sentences or sentence fragments using the words "ball," "base," and "bat" together or in similar contexts, the vector numbers will also reflect this distinct meaning of the word "bat" by reflecting its semantic proximity to this different set of words. For illustrative purposes only, below is an example of what the resulting vectors might look like[10]:

watch	2.6	2.6	2.9	1.5	2.5	2.0	0.8	1.2
base	1.3	1.0	0.2	1.4	4.3	2.3	3.3	2.3
ball	1.2	2.1	0.1	3.4	1.0	2.4	3.4	4.4
bat	1.1	3.4	0.1	0.3	1.7	1.6	3.5	5.5
night	0.2	3.5	4.0	0.3	1.8	3.4	1.2	0.2
fly	0.5	3.3	3.4	0.2	2.3	4.2	2.0	1.3
emily	1.6	0.7	1.7	1.6	1.0	2.0	0.2	0.8

As shown above, the semantic proximity of the words "base," "ball," and "bat" is represented by the proximity of the numbers in the first, third, and seventh vector spaces (highlighted in gray). And the semantic proximity of the words "bat," "night," and "fly" is represented by the numbers in the second and fourth vector spaces (highlighted in blue). In this way, the model's vector for the word "bat" simultaneously reflects both possible meanings of that word. After examining billions of similar sentences — and making billions upon billions of tiny adjustments to the numbers in these vectors — the word vectors eventually become a semantic map of the language, accurately reflecting the meaning of words by encoding the semantic relationship between them.

[10] *See generally* Meta, Comments in Response to the Copyright Office's Notice of Inquiry on Artificial Intelligence and Copyright at *5 (Oct. 30, 2023), https://www.regulations.gov/comment/COLC-2023-0006-9027.

Next, the raw model is further optimized to follow instructions and improve performance on specific tasks, where it is exposed to more specialized datasets to improve performance on specific tasks.[11] Numerous models undergo Reinforcement Learning from Human Feedback (RLHF), where human reviewers rank generated responses, helping the model learn what makes a response useful and appropriate.[12] These ranked responses are used to train a separate neural network — the reward model — which learns to assign a numerical "reward score" to any given model output. The higher the score, the more likely the response aligns with human preferences.[13]

Because the training data is broken down into numerical representations and embedded in a complex network of parameters, its raw form need not be stored or systematically retrieved for generation of outputs. As discussed in section II(A) below, the embeddings reflect patterns, structures, and relationships in the aggregate, from across the vast training dataset as a whole, not any individual work.

II. Embedded Statistical Relationships in AI Models Do Not Constitute "Copies" or "Derivatives" of AI Training Data

Copyright is a form of protection that U.S. law extends to "original works of authorship" that are "fixed in any tangible medium."[14] Copyright owners are granted the exclusive rights of reproduction (copying) and preparing derivative works (works based upon one or more preexisting works), among others.[15] The primary objective of copyright, however, is not "to reward the labor of authors."[16] Rather, the goal of copyright is "'[t]o promote the Progress of Science and useful Arts'"[17] and "to expand public knowledge and understanding."[18] "To this end, copyright assures authors the right to their original expression, but encourages others to build freely upon the ideas and information conveyed by a work."[19]

[11] Venkatesh Balavadhani Parthasarathy et al., *The Ultimate Guide to Fine-Tuning LLMs from Basics to Breakthroughs: An Exhaustive Review of Technologies, Research, Best Practices, Applied Research Challenges and Opportunities*, ARXIV (Oct. 30, 2024), https://arxiv.org/pdf/2408.13296.

[12] Long Ouyang et al., *Training Language Models to Follow Instructions with Human Feedback*, ARXIV (Mar. 4, 2022), https://arxiv.org/pdf/2203.02155.

[13] *Id.*

[14] 17 U.S.C. § 102(a).

[15] 17 U.S.C. § 106.

[16] *Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., Inc.*, 499 U.S. 340, 349 (1991).

[17] *Id.* (citing U.S. Const. art. I, § 8, cl. 8).

[18] *Authors Guild v. Google, Inc.*, 804 F.3d 202, 212 (2d Cir. 2015) ("*Google Books*").

[19] *Feist Publ'ns*, 499 U.S. at 349–50.

For this reason, although copyright law protects creative expression, it does not extend to the underlying ideas, theories, and facts expressed in a work[20]; those "become[] instantly available for public exploitation at the moment of publication."[21] Accordingly, while copyright grants owners certain exclusive rights, copyright was never intended to prohibit learning from or being inspired by creative works to create new ones. An increase in ideas and competitive creative works is, in fact, precisely the growth of creative expression that the Copyright Act was intended to promote.[22]

In the context of AI, as discussed above, the technical process of "learning" for AI models means being able to derive patterns, structures, and relationships from across a broad body of training data so that the models can generate appropriate responses to user prompts based on statistical relationships. AI models are not designed to store or systematically retrieve content. The training data is extensively transformed, processed, and then broken down into numerical representations of the vast training corpus as a whole and embedded in a complex network of parameters. It is these patterns — not the training data's protectable expression — that are retained by the model. Models use the data to develop an entirely new and innovative service that, in turn, produces valuable *new* content — thereby vastly expanding the capacity for human creative productivity.

Nevertheless, some commentators have suggested that the use of copyrighted works for AI training purposes could constitute copyright infringement because numerical representations of the training data remain within the AI model. They maintain that because a model — when intentionally prompted — can generate a verbatim or substantially similar copy of a training example, a copy of that example must exist in some form in the model's weights and may therefore infringe the derivative work right.

In the next few sections, we address multiple mischaracterizations about the embeddings retained in AI models to show that embedded statistical relationships do *not* constitute "copies" or "derivatives" of AI training data. In addition, in section III, we discuss how to the extent any copies of copyrighted works are made during AI training, such copying constitutes fair use under existing law.

[20] 17 U.S.C. § 102(b); *see also Google Books*, 804 F.3d at 225.

[21] *Eldred v. Ashcroft*, 537 U.S. 186, 219 (2003); 17 U.S.C. § 102(b) ("In no case does copyright protection for an original work of authorship extend to any idea, procedure, process, system, method of operation, concept, principle, or discovery, regardless of the form in which it is described, explained, illustrated, or embodied in such work.").

[22] *Sega Enters. Ltd. v. Accolade, Inc.*, 977 F.2d 1510, 1523 (9th Cir. 1992).

A. Embeddings Reflect Statistical Relationships Learned from the Training Dataset as a Whole, Not Specific Inputs

AI models do not "decompile" or "reverse" embeddings back into original sentences or training data. For example, when generating responses, LLMs predict (e.g., one word at a time) based on learned statistical patterns and construct original responses rather than retrieve content. Moreover, embeddings encode statistical relationships rather than semantic meanings from text and are generalized *across the dataset in its entirety*, creating representations that are not tied to any specific work within a training dataset. The purpose of embedding is to represent the full set of semantic meanings for a particular token, as seen across all training data — not to reflect a single use of a term in a single copyrighted work. Ultimately, LLMs generate text probabilistically, meaning they select words based on statistical distributions, and are non-deterministic (i.e., the same user prompts often produce different outputs).[23]

The ability to recombine learned language and structures rather than merely regurgitate training data is what makes AI models effective in generating diverse and contextually appropriate responses.

For example, if a user prompt states, "Do queens live in a castle?", the model references the vector representations of key words like "queen" and "castle" to identify their contextual relationships. Because "queen" is mapped closely to words like "king" and "royal" in its learned vector space, and "castle" is strongly associated with "monarchs" and "residences," the model recognizes that these concepts frequently co-occur. Using this learned pattern, the model predicts that the most probable response would affirm this relationship, and the predicted token outputs are arranged in a manner based on how these words have co-occurred in its training data as a whole. As a result, the model might generate: "Yes, as royals, kings and queens often live in castles." Although such a response may resemble common sentences in the training data, it is not a retrieval of a string of text, but rather a new construction based on word associations.

Even if a model could "look up" these vectors, the model would likely produce an imprecise reconstruction or paraphrase. If the sentence "The princesses had an unbelievable victory in yesterday's game" appeared in training data and the model theoretically could "look up" these vectors, the model might generate a sentence like "The princesses had a not belief victory in yesterday game," an imprecise reconstruction that loses the original phrasing. As discussed in the next section, even if verbatim "re-presentation" should occur, it is not indicative of systematic or widespread "lossless" copying, as outputs are designed to be probabilistic reconstructions influenced by a wide variety of training inputs.

[23] Lokke Moerel & Marijn Storm, *Do LLMs "Store" Personal Data? This Is Asking the Wrong Question*, IAPP (Oct. 23, 2024), https://iapp.org/news/a/do-llms-store-personal-data-this-is-asking-the-wrong-question/.

B. Verbatim Reconstruction of Training Data May Result from Imprecise Training Practices

Although a language model does not keep an index of books, articles, or websites, it does retain the statistical correlations of its training dataset, which are used to produce new, not stored, sentences. There are instances, however, in which the model may output a sequence of text that closely resembles, or even exactly matches, a portion of its training data, a phenomenon often referred to as "memorization" or "regurgitation." Regurgitation typically occurs when there are flaws in the training pipeline, including (1) oversampling: the same excerpt appears so many times that the model assigns it an outsized statistical weight, (2) deduplication failure: identical or near-identical strings are not filtered out before training that the string receives excessive statistical weight, and/or (3) lack of diversity in training data: distinctive or uncommon phrases in a non-diverse dataset can lead to the model assigning the distinctive phrase disproportionate weight which may lead to regurgitation, especially when the user prompt closely mirrors the training context of the distinctive phrase. Thus, while it may appear in some cases that the model has "stored" the original content, such behavior is more accurately understood as a byproduct of imprecise training practices. Such cases do not reflect the standard operation of a properly trained model.[24]

Indeed, to reduce the chances of generating even trivial "re-presentation," AI developers employ several techniques.[25] For example, dataset deduplication serves to remove duplicate text before training begins, preventing overrepresentation of any single passage. Also, "attention mechanisms" are used to dynamically focus the model on the parts of a user's prompt input that are most relevant, allowing the model to generate tailored responses.[26] In addition, context analysis ensures that previous words in a conversation influence the response, allowing the model to generate answers that are tailored to the specific user prompt rather than simply producing a generic output. Also, temperature settings regulate how creative or predictable the output is by adjusting the randomness in word selection so that responses can be more varied and unexpected as opposed to producing the most common output. Finally, reinforcement learning tunes models to prioritize generalization over direct reproductions, guiding them toward generating contextually relevant yet original responses.

---

[24] Tong Chen et al., *ParaPO: Aligning Language Models to Reduce Verbatim Reproduction of Pre-training Data*, ARXIV (APR. 20, 2025), https://arxiv.org/abs/2504.14452.

[25] Chiyuan Zhang et al., *Counterfactual Memorization in Neural Language Models*, ARXIV (Oct. 13, 2023), https://arxiv.org/pdf/2112.12938.

[26] Ashish Vaswani et al., *Attention Is All You Need*, ARXIV (Aug. 2, 2023), https://arxiv.org/abs/1706.03762.

C. Statistical Relationships Reflected in Embeddings Are Based on Unprotectable Ideas, Not Expressive Content

While copyright law protects creative expression, it does not extend to the underlying ideas, theories, and facts expressed in a work. Accordingly, a copyright owner's rights to reproduce and prepare derivative works extend only to the expressive content within an original work, not to any factual information or concepts reflected in or about the work.[27]

Using statistical information about a work — such as "word frequencies, syntactic patterns, and thematic markers [] to derive information on [] nomenclature, linguistic usage, and literary style" — has been found *not* to implicate the interests protected by copyright.[28] The same principle applies to AI model embeddings, as they provide information about the training data so the models can derive patterns, structures, and relationships from across a broad body of content and be able to spontaneously generate outputs based on the statistical patterns learned during training.

Moreover, incorporating ideas and concepts from one work into another does not create a "derivative" work in violation of the copyright owner's exclusive rights. A "derivative work" is a "work based upon one or more preexisting works …, or any other form in which a work may be recast, transformed, or adapted."[29] A derivative work, however, "does not refer to all works that borrow in any degree from pre-existing works."[30] "If what is borrowed consists *merely of ideas* and not of the expression of ideas, then, although the work may have in part been derived from prior works, it is not a derivative work."[31] To the extent commentators have suggested that AI models are themselves infringing derivative works of their training data, that notion has been rejected as "nonsensical."[32] For a new work to be an infringing derivative, it must be "substantially similar" to the protectable expressive elements of the original work.[33]

---

[27] *Google Books*, 804 F.3d at 207, 225.

[28] *See id.* at 225–26 ("The copyright resulting from the Plaintiffs' authorship of their works does not include an exclusive right to furnish … information about the works.").

[29] 17 U.S.C. § 101.

[30] 1 MELVILLE B. NIMMER & DAVID NIMMER, NIMMER ON COPYRIGHT § 3.01 (2025).

[31] *Id.* (emphasis added).

[32] *See Kadrey v. Meta Platforms Inc.*, No. 23-CV-03417, 2023 U.S. Dist. LEXIS 207683, at *1 (N.D. Cal. Nov. 20, 2023).

[33] *See id.* ("To prevail on a theory that [an LLM's] outputs constitute derivative infringement, the plaintiffs would indeed need to allege and ultimately prove that the outputs 'incorporate in some form a portion of' the plaintiffs' books.").

III. Consistent with Legal Precedent, the Process of Training an AI Model Constitutes Fair Use

A. As Intended, Fair Use Provides a "Check" on the Copyright Monopoly

"[A] copyright holder cannot prevent another person from making a 'fair use' of copyrighted material."[34] The doctrine of fair use permits copying and other uses of copyrighted works without permission where — as with training AI models — the secondary use is for a new (i.e., "transformative") purpose and does not function as a market substitute for the original copyrighted works.

As was Congress's goal, fair use has played a crucial role in promoting American innovation and technological developments. Congress recognized fair use as "one of the most important and well-established limitations on the exclusive right of copyright owners," which should be "adapt[ed]" to account for "rapid technological change."[35] Congress noted that fair use had "been raised as a defense in innumerable copyright actions over the years," and that there had been "ample case law recognizing the existence of the doctrine and applying it."[36]

When Congress codified the common-law doctrine of fair use in the 1976 Copyright Act, the doctrine had already been an essential part of U.S. copyright law for more than 130 years.[37] Now, nearly 185 years after fair use was first considered by a U.S. court, section 107 of title 17 of the United States Code provides: "[T]he fair use of a copyrighted work, … for purposes such as criticism, comment, news reporting, teaching … scholarship, or research, is not an infringement of copyright." To determine whether a particular use is "fair," the statute provides four factors to be considered:

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.[38]

---

[34] *Google LLC v. Oracle Am., Inc.*, 593 U.S. 1, 18 (2021) (citing 17 U.S.C. § 107).

[35] H.R. REP. No. 94–1476 at 65–66 (1976).

[36] H.R. REP. No. 94–1476 at 65 (1976).

[37] *See Folsom v. Marsh*, 9 F. Cas. 342, 348 (C.C.D. Mass. 1841) ("In short, we must often … look to the nature and objects of the selections made, the quantity and value of the materials used, and the degree in which the use may prejudice the sale, or diminish the profits, or supersede the objects, of the original work.").

[38] 17 U.S.C. § 107.

Time and time again, courts have found fair uses of copyrighted works to prevent copyright from exceeding "its lawful bounds" and obstructing the development and distribution of new and innovative technologies.[39] As discussed below under the third fair-use factor, even in cases involving *full* copying and retention of copyrighted works, courts have found such uses to be fair where they further the development of new and innovative technologies.

Below, we apply the fair use factors to show that, to the extent any reproductions of copyrighted materials are made during AI model training, such use qualifies as a transformative use and does not function as a market substitute for the original works, thus making such use "fair."

B. Factor One Weighs in Favor of Fair Use

Under the first fair use factor — the purpose and character of the secondary use — using copyrighted works to train AI models is a "transformative" use that serves a different purpose from that of the original work.

According to the Supreme Court in *Warhol v. Goldsmith*, the "central" question for this factor is where the new use is "transformative" — that is, "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."[40] "[A] use that has a distinct purpose is justified" and is thus fair "because it furthers the goal of copyright, namely, to promote the progress of science and the arts, without diminishing the incentive to create."[41] While commerciality of the secondary use may be relevant, it is not dispositive.[42]

---

[39] *See, e.g.*, *Google*, 593 U.S. at 31 (finding the use of declarations copied from a computer program's application programming interface (API) to be fair where it "further[ed] the development of computer programs" by allowing programmers to use their acquired skills to develop new applications for a new platform); *Google Books*, 804 F.3d at 217 (finding fair use where the digitizing of millions of books to enable a search function was transformative and not meant as a substitute for the authors' books); *Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87, 97 (2d Cir. 2014) (similar; "the result of a word search is different in purpose, character, expression, meaning, and message from the page (and the book) from which it is drawn"); *Sega Enters.*, 977 F.2d at 1522 (finding fair use where computer code was copied for the purpose of reverse engineering and studying how to develop new video games that were compatible with an existing game console).

[40] *Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith*, 598 U.S. 508, 528 (2023) ("*Warhol*") (citing *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 579 (1994)).

[41] *Id.* at 531 (citing *Campbell*, 510 U.S. at 579; *Google Books*, 804 F.3d at 214).

[42] *Warhol*, 598 U.S. at 531.

In reaching its determination in *Warhol*, the Supreme Court approvingly cited *Google Books* three times, a case in which the Second Circuit held that copying and digitizing millions of books to enable a search function "involve[d] a highly transformative purpose" and was not meant as a substitute for the authors' books.[43] The court found that the purpose of the copying was "to make available significant information *about those books*," permitting users to research "word frequencies, syntactic patterns, and thematic markers" and "derive information on … nomenclature, linguistic usage, and literary style."[44]

Significantly, the first U.S. court to issue a substantive decision regarding fair use and generative AI model training relied on *Warhol* and *Google Books*, holding in *Bartz v. Anthropic* that the use of copyrighted works to train large language models is "exceedingly transformative."[45] Two days later, in *Kadrey v. Meta Platforms, Inc.*, a second U.S. court determined that an AI company's use of books for LLM training "had a 'further purpose' and 'different character' than the books — that it was highly transformative."[46]

These courts reached the correct conclusion. As the court held in *Bartz*, deriving patterns, structures, and relationships *about the content*, across a vast dataset, for the purpose of creating an AI model with the ability to create *new* content, is "quintessentially transformative."[47] Finding transformativeness where the copying is done to discern functional patterns, structures, and relationships is consistent with *Google Books* and other legal precedent.[48]

In *Bartz*, the court also concluded that, under the first factor, AI model training does not supplant copyrighted works used in training data.[49] The court noted that the AI model "trained upon works not to race ahead and replicate or supplant them — but to turn a hard corner and create something different."[50] The court made the right determination. Training data is used to develop an entirely new and innovative service that, in turn, produces valuable *new* content — thereby vastly expanding the capacity for human creative productivity.

---

[43] *Google Books*, 804 F.3d at 216.

[44] *Google Books*, 804 F.3d at 209, 217; *see also HathiTrust*, 755 F.3d at 97 (finding that copying millions of books to create a searchable database was "quintessentially transformative" because "the result of a word search is different in purpose, character, expression, meaning, and message from the page (and the book) from which it is drawn").

[45] *Bartz v. Anthropic PBC*, No. C 24-05417, 2025 WL 1741691, at *5 (N.D. Cal. June 23, 2025).

[46] *Kadrey*, 2025 WL 1752484, at *9.

[47] *Bartz*, 2025 WL 1741691, at *8.

[48] *See, e.g.*, *Google Books*, 804 F.3d at 209, 217; *Sega Enters.*, 977 F.2d at 1522 (finding transformative use where computer code was copied for the purposes of reverse engineering — to discern the functional patterns, structures, and relationships necessary to understand the software's operation — to develop new video games that were compatible with an existing game console); *Sony Computer Entm't, Inc. v. Connectix Corp.*, 203 F.3d 596, 606–07 (9th Cir. 2000) (finding transformative use where computer code was copied for the purposes of reverse engineering to create a "new platform" on which to play existing videogames).

[49] *Bartz*, 2025 WL 1741691, at *8.

[50] *Id.*

Some commentators have asserted that copyrighted works are used for AI model training due to their expressive and aesthetic value, and so AI training shares the same "purpose" of the training data. The court in *Kadrey* rejected such an argument, however.[51] The court disagreed that the AI company's use had the same purpose and character as the books because an LLM training on a book is akin to a human reading one. Rather, "an LLM's consumption of a book is different than a person's, as an LLM ingests text to learn 'statistical patterns' of how words are used together in different contexts, whereas this is not how a human reads a book."[52] Using copyrighted works to learn textual patterns, structures, and relationships "add[s] something new, with a further purpose or different character."[53] Consequently, the use of copyrighted works for AI training does not "supersede[] the objects [or purposes] of the original creation."[54]

Other commentators have incorrectly asserted that, under *Warhol*, finding transformativeness in the context of AI model training will depend on the functionality of the model and how it is deployed. Such an approach, however, places undue emphasis on potential outputs generated by a model. Whether the AI model generates similar expressive *outputs* is not relevant in determining whether *AI training* is transformative. Rather, the evaluation of outputs is relevant to whether the *outputs* are infringing of (i.e., substantially similar to) prior works. To conclude otherwise would hold AI developers responsible for unknowable, post-training activity by third parties. Evaluating training through the lens of outputs improperly conflates two distinct issues: whether the use of data for training was lawful, and how the model is later deployed.

Improperly focusing on a model's outputs is also in tension with *Warhol* and legal precedent determining that the fair use analysis should focus on the specific allegedly infringing activity committed by the specific alleged infringer. In *Warhol*, the Supreme Court evaluated fair use by focusing on whether the 2016 licensing of a small-scale copy of Warhol's previously created large-scale screenprint was infringing — the particular use being challenged in the case — not the original screenprint's creation decades earlier.[55]

In sum, for AI model training, the use of copyrighted works is to create a model capable of producing *new* expressive content when deployed in a generative AI system. Such a purpose is "quintessentially transformative."[56]

[51] *Kadrey*, 2025 WL 1752484, at *10.
[52] *Id.*
[53] *Warhol*, 598 U.S. at 528 (citing *Campbell*, 510 U.S. at 579); *see Google Books*, 804 F.3d at 209.
[54] *HathiTrust*, 755 F.3d at 97 (citing *Campbell*, 510 U.S. at 579) (internal quotation marks omitted).
[55] *Warhol*, 598 U.S. at 534.
[56] *Bartz*, 2025 WL 1741691, at *8.

C. Factor Two Rarely Plays a Role in the Fair Use Analysis

The second factor — the nature of the copyrighted work — reflects the fact that not all copyrighted works are entitled to the same level of protection.[57] In considering the nature of the copyrighted work, the Supreme Court has instructed that "fair use is more likely to be found in factual works than in fictional works," whereas "a use is less likely to be deemed fair when the copyrighted work is a creative product."[58]

While part of the overall fair use analysis, the second factor "has rarely played a significant role in the determination of a fair use dispute."[59] Accordingly, courts may give this factor little weight in the overall fair use analysis, especially when the secondary use is transformative (as with AI model training) or when the other factors strongly point one way or the other.[60]

D. Factor Three Weighs in Favor of Fair Use

According to the Supreme Court, the third fair use factor focuses, in particular, on whether "the amount and substantiality of the portion used … [is] reasonable in relation to the purpose of the copying."[61] "Complete unchanged copying has repeatedly been found justified as fair use when the copying was reasonably appropriate to achieve the copier's transformative purpose and was done in such a manner that it did not offer a competing substitute for the original."[62] In other words, even full copying and retention of copyrighted works does not prevent a use from being "fair."[63]

In the context of evaluating the use of copyrighted works to train LLMs, the courts in *Bartz* and *Kadrey* both determined that full copying of training data is "reasonably necessary" and thus weighs in favor of fair use.[64] As noted by the court in *Kadrey*, "feeding a whole book to an LLM does more to train it than would feeding it only half of that book."[65]

---

[57] *Sega Enters.*, 977 F.2d at 1524.
[58] *Stewart v. Abend*, 495 U.S. 207, 237 (internal quotation marks and alteration omitted).
[59] *See Google Books*, 804 F.3d at 220 (citing WILLIAM F. PATRY, PATRY ON FAIR USE § 4.1 (2015)).
[60] *See, e.g.*, *Campbell*, 510 U.S. at 586 (finding second factor was "not much help"); *HathiTrust*, 755 F.3d at 98 (finding second factor "is not dispositive").
[61] *Campbell*, 510 U.S. at 586.
[62] *Google Books*, 804 F.3d at 221.
[63] *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417, 449–50 (1984).
[64] *Kadrey*, 2025 WL 1752484, at *14; *Bartz*, 2025 WL 1741691, at *15–16.
[65] *Kadrey*, 2025 WL 1752484, at *14.

Such a conclusion is consistent with case law finding that copying an entire work is fair where reasonable or necessary to achieve the purpose of the fair use. For example, in *Google Books*, the court held that "not only is the copying of the totality of the original [works] reasonably appropriate to [the] transformative purpose" — creating a tool to search for and return snippets of books containing certain words — "it is literally *necessary* to achieve that purpose."[66] Similarly, in *Author's Guild v. HathiTrust*, the Second Circuit found that copying millions of books in their entirety to create a searchable database "was reasonably necessary," as it "enable[d] the full-text search function" for the database and was thus not excessive.[67] In *Kelly v. Arriba*[68] and *Perfect 10 v. Amazon.com*,[69] the Ninth Circuit held that the defendants' wholesale copying of images was necessary to generate smaller, lower-resolution "thumbnail" images for display on search results pages. The Ninth Circuit has also found fair use where entire software code was copied to access its unprotected functional elements for reverse engineering purposes.[70] And in *A.V. ex rel. Vanderhye v. iParadigms, LLC*, the Fourth Circuit held that the complete copying of unaltered student papers did not preclude a finding of fair use where the copies were used in connection with a computer program that detects plagiarism.[71]

E. Factor Four Weighs in Favor of Fair Use

The fourth fair use factor — "the effect of the use upon the potential market for or value of the copyrighted work" — focuses on whether the secondary use "usurps the market of the original work,"[72] or, in other words, "brings to the marketplace a competing substitute for the original."[73] The goal of the fourth factor is not, however, to protect copyright owners from *any* form of "economic harm."[74] As noted by the Supreme Court, "a potential loss of revenue is not the whole story," as a "lethal parody, like a scathing theatre review," may "kil[l] demand for the original."[75] Such harm, "even if directly translated into foregone dollars, is not 'cognizable under the Copyright Act.'"[76]

---

[66] *Google Books*, 804 F.3d at 221 (emphasis added).

[67] *HathiTrust*, 755 F.3d at 98. The court also found that the retention of digital image files and text-only files in the library was necessary to provide access for individuals with disabilities, as the text files were required for text searching and to create text-to-speech capabilities and the image files were necessary to perceive the books fully. *HathiTrust*, 755 F.3d at 102–03.

[68] *Kelly v. Arriba Soft Corp.*, 336 F.3d 811, 821 (9th Cir. 2003). The court determined that full copying was reasonable "to allow users to recognize the image and decide whether to pursue more information about the image or the originating [website]." *Id.*

[69] *Perfect 10 v. Amazon.com, Inc.,* 508 F.3d 1146, 1167–68 (9th Cir. 2007).

[70] *See, e.g.*, *Sony Comput.*, 203 F.3d at 606, 610; *Sega Enters.*, 977 F.2d at 1527–28.

[71] *A.V. ex rel. Vanderhye v. iParadigms, LLC*, 562 F.3d 630, 634, 642 (4th Cir. 2009).

[72] *HathiTrust*, 755 F.3d at 99.

[73] *Google Books*, 804 F.3d at 223.

[74] *HathiTrust*, 755 F.3d at 99.

[75] *Google*, 593 U.S. at 13.

[76] *Id.* (citing *Campbell*, 510 U.S. at 591–92).

Significantly, where the transformative purpose of the secondary use is not designed to result in "widespread revelation of sufficiently significant portions of the original," the use does not "make available a significantly competing substitute."[77] Here, AI models are not regurgitating training data and making competing substitutes; rather, they are creating *new* content in response to user prompts. Where an AI model does not generate any "meaningful portion" or "exact copies" of training data, courts have found no "meaningful or significant effect 'upon the potential market for or value of'" copyrighted works used for AI training.[78]

Some commentators have asserted that because certain AI model developers have entered data use agreements with content owners, this factor may weigh against fair use. Courts, however, have disagreed with such a position. Whether a licensing market for AI training exists or is likely to develop is "irrelevant," as "this market is not one that [copyright owners] are legally entitled to monopolize."[79] Otherwise, there is "a danger of circularity" under this factor, as "it is always a given that [the] plaintiff suffers a loss of *some* potential market if that potential is defined as the theoretical market for licensing the very use at bar."[80] To "prevent the fourth factor analysis from becoming circular and favoring the rightsholder in every case, harm from the loss of fees paid to license a work for a transformative purpose is not cognizable."[81]

---

[77] *Google Books*, 804 F.3d at 223–24 (finding that the "snippet" view produced only "discontinuous, tiny fragments" of the books being searched, amounting in the aggregate to a small percentage of the books' expressive content, which did "not threaten the rights holders with any significant harm to the value of their copyrights or diminish their harvest of copyright revenue").

[78] *Kadrey*, 2025 WL 1752484, at *15 (finding no market harm where an LLM "does not allow users to generate any meaningful portion of the plaintiffs' books"); *Bartz*, 2025 WL 1741691, at *8 (similar; "Authors concede that training LLMs did not result in any exact copies nor even infringing knockoffs of their works being provided to the public.").

[79] *Kadrey*, 2025 WL 1752484, at *16.

[80] 1 MELVILLE B. NIMMER & DAVID NIMMER, NIMMER ON COPYRIGHT § 13F.08[B] (2025).

[81] *Kadrey*, 2025 WL 1752484, at *16; *Bartz*, 2025 WL 1741691, at *17 (rejecting argument that training LLMs displaced (or will displace) an emerging market for licensing their works for the narrow purpose of training LLMs, as "such a market for that use is not one the Copyright Act entitles Authors to exploit"); *see also HathiTrust*, 755 F.3d at 99–100 (finding that because the full-text search function did not serve as a substitute for the books being searched, "it [was] irrelevant that the Libraries might be willing to purchase licenses in order to engage in th[e] transformative use").

Another theory of market harm recently suggested is one of "market dilution." In *Kadrey*, the court stated that another way "using copyrighted books to train an LLM might harm the market for those works is by helping to enable the rapid generation of countless works that compete with the originals."[82] Despite the plaintiffs "present[ing] *no evidence* about how the current or expected outputs … would dilute the market for their own works" — and not even raising market dilution as an issue in their Complaint or their own motion for summary judgment — the court maintained that this form of competition could "dilute" the market for copyrighted books, even if the AI-generated works "aren't themselves infringing."[83]

The court in *Bartz*, however, correctly rejected this "market dilution" theory. Such a theory runs counter to copyright law's core goal of encouraging new creative works and extends far beyond the traditional understanding of how courts have applied the fourth fair use factor. Copyright law protects works, not markets. As held in *Bartz*, the creation of AI-generated competing expressive works "is not the kind of competitive or creative displacement that concerns the Copyright Act," which "seeks to advance original works of authorship, not to protect authors against competition."[84] If a new work does not use protected expression, it does not matter whether it competes in the same genre and market as prior works. An increase in competitive creative works is precisely the growth of creative expression that the Copyright Act was intended to promote.[85] Indeed, the ability of a new technology to spur the creation of new expressive works is a public benefit the Supreme Court has said should be considered when evaluating this factor.[86]

[82] *Kadrey*, 2025 WL 1752484, at *16.

[83] *Id.* at *2, *16 (emphasis added). The court acknowledged but could not reconcile, however, that "an LLM trained *only on public domain works* could still be capable of quickly generating large numbers of books that could compete for sales with copyrighted books." *Id.* at *17 (emphasis added).

[84] *Bartz*, 2025 WL 1741691, at *17.

[85] *Sega Enters.*, 977 F.2d at 1523.

[86] *Google*, 593 U.S. at 35 ("[W]e must take into account the public benefits the copying will likely produce. Are those benefits, for example, related to copyright's concern for the creative production of new expression?").

Moreover, attempts to prove "market dilution" would likely be too attenuated and speculative for thorough evaluation under the fourth factor. Numerous factors — market trends, economic conditions, consumer preferences, and human-authored competing works — can influence sales rather than the mere existence of AI-generated works in the marketplace. For market harm to be cognizable, it must be concrete and probable — not merely possible. Courts have rejected assertions of market harm under the fourth fair use factor where the potential future harm is merely speculative.[87] Predicting that consumers will substitute AI-generated books for human-authored ones, merely because they may exist, is inherently uncertain and should not serve as the basis for finding against fair use.

IV. Conclusion

In sum, the transformative use of copyrighted works for AI model training furthers the constitutional goal of "promot[ing] the Progress of Science and useful Arts"[88] by enabling the development and distribution of groundbreaking technologies that enable the creation of new expressive content. To that end, it is vital to resist efforts that would impose undue regulatory burdens on AI model developers in this vital stage of innovation. Introducing new statutory requirements or restrictions would create unnecessary barriers to progress, hinder economic growth, and weaken America's competitive edge in the global AI landscape.

Given the profound implications for the future of AI, it is critical to recognize that the U.S. doctrine of fair use is an essential component of any legal framework that will govern AI, and that the use of copyrighted materials to train AI models is a fair use as a matter of law.

If the United States is to remain the global leader in AI, its legal framework must continue to support both creativity and innovation, not put them at odds.

---

[87] *See, e.g.*, *HathiTrust*, 755 F.3d at 100–01 (rejecting assertion of potential harm from hackers obtaining unauthorized access to the books stored online, seeing "no basis in the record on which to conclude that a security breach is likely to occur") (citing *Universal City Studios*, 464 U.S. at 453–54 (concluding that time-shifting using a VCR is fair use because the copyright owners' "prediction that live television or movie audiences will decrease" was merely "speculative") (comparing *Clapper v. Amnesty Int'l*, 568 U.S. 398, 401 (2013) (finding risk of future harm must be "certainly impending," rather than merely "conjectural" or "hypothetical," to constitute a cognizable injury-in-fact)); *Google Books*, 804 F.3d at 227 (similar); *see also Perfect 10 v. Amazon.com, Inc.,* 508 F.3d 1146, 1168 (9th Cir. 2007) (finding "potential harm" to market remained "hypothetical").

[88] U.S. Const. art. I, § 8, cl. 8.


AI
PROGRESS