Testimony before the U.S. Senate Committee on the Judiciary
Subcommittee on Crime and Counterterrorism

Hearing on Too Big to Prosecute?: Examining the AI Industry's Mass Ingestion of Copyrighted Works for AI Training

July 16, 2025

Edward Lee
Professor of Law
Santa Clara University School of Law

Chair Hawley, Ranking Member Durbin, and other Members of the Subcommittee: Thank you for the opportunity to testify. I am a Professor of Law at Santa Clara University School of Law.[1] I am also a book author and a photographer,[2] and my personal experience informs my scholarship and understanding of the importance of copyright to authors and artists across the country.

In my testimony, I will discuss whether using copyrighted works to train AI models is a fair use, giving particular attention to the two recent decisions by Judges Alsup and Chhabria in cases filed by book authors against Anthropic and Meta. This novel question of law, which has important implications for U.S. national interest, has sparked sharp disagreements among parties, stakeholders, and now federal judges. As Judge Bibas noted in an earlier non-generative AI case, this question of law is difficult.[3] In my opening remarks, I would like to stress three points.

Transformative purpose of AI training. First, I believe Judges Alsup and Chhabria correctly concluded that the use of copies of works to train an AI model serves a highly transformative purpose in developing a new technology under Factor 1 of fair use.[4] During training, an AI model is exposed to vast training materials, typically many millions of works. Through a process called deep learning, the model identifies the "statistical relationships among words," thereby enabling the model to conduct numerous functions, including research, translation of foreign languages, delivery of medical advice, generation of content, and so forth.[5] As Judge Chhabria concluded, "The purpose of Meta's copying was to train its LLMs [large language models], which are innovative tools that can be used to generate diverse text and perform a wide range of

[1] My law review article on "Fair Use and the Origin of AI Training" will be published by Houston Law Review. *See* Edward Lee, *Fair Use and the Origin of AI Training*, 63 HOU. L. REV. (forthcoming 2025), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5253011 [hereinafter *Origin of AI Training*]. I published other articles on copyright issues raised by generative AI. *See* Edward Lee, *AI and the Sound of Music*, 134 YALE L. J. FORUM 187 (2024); Edward Lee, *Prompting Progress: Authorship in the Age of AI*, 76 FLA. L. REV. 1445 (2024). On my website CHATGPT IS EATING THE WORLD, I track and analyze all the U.S. copyright lawsuits—currently 44 pending lawsuits—against AI companies. I have included a map of the United States listing these cases attached at the end of this statement as Appendix E.
[2] *See* Edward Lee, CREATORS TAKE CONTROL (2023); Edward Lee, PICERRIFIC PHOTOGRAPHY.
[3] Thomson Reuters Enter. Centre GMBH v. ROSS Intell., Inc., 2025 WL 1488015, at *1 (D. Del. May 23, 2025) (granting petition to file interlocutory appeal on fair use and copyrightability of headnotes while noting "[o]ur circuit has not yet spoken on this 'novel and difficult question[ ] of first impression.'") (internal citation omitted); *id.* ("these questions are hard").
[4] *See* Bartz v. Anthropic PBC, -- F. Supp. 3d --, 2025 WL 1741691, at *7 (N.D. Cal. Jun. 23, 2025); Kadrey v. Meta Platforms, Inc., -- F. Supp. 3d --, 2025 WL 1752484, at *9 (N.D. Cal. June 25, 2025). I summarize the decisions in attached Appendix A.
[4] *Kadrey*, 2025 WL 1741691, at *9-*10. The four factors of fair use in Section 107 are quoted in Appendix A.
[5] *Id.* at *5, *9.

functions."[6] And, as Judge Alsup recognized, "The technology at issue was among the most transformative many of us will see in our lifetimes."[7]

The history of AI development strongly supports this conclusion. It is important to understand *why* AI researchers at universities began training AI models on large datasets. The practice originated, not at AI companies, but at universities where AI researchers discovered a key insight: *scaling*, or using larger and more diverse datasets actually worked in developing and improving AI models—an achievement that escaped researchers for many years.[8] This seminal breakthrough, which took decades to figure out, propelled the advances in AI witnessed today.

Some uses might not be fair. Second, while I agree with the ultimate findings of fair use in both cases, it's important to remember that fair use is a fact-specific doctrine decided on a case-by-case basis. In some situations, a transformative purpose in AI training might be outweighed by other fair use factors. For example, an AI model that routinely produces outputs that are infringing, such as regurgitations, might not be a fair use—even in the training—due to insufficient guardrails. Critically, in the cases against Anthropic and Meta, the plaintiffs did *not* show the models produced infringing outputs of their works.[9]

National priority in AI innovation. My final point is the need for caution—caution by the courts, Congress, and the states. I believe it's important to weigh the United States' interest in AI innovation. President Trump issued an executive order making U.S. development and global leadership in AI a national priority.[10] China has its own priority and a plan—of surpassing the United States and becoming the world leader in AI by 2030.[11] The United States' national priority in AI counsels caution.

Indeed, in *Google v. Oracle*, another technology fair use case of national importance, the Supreme Court itself cautioned: "Given the rapidly changing technological, economic, and business-related circumstances, we believe we should not answer more than is necessary to resolve the parties' dispute."[12] Judges Alsup and Chhabria departed from this approach in some controversial parts of their opinions that were just dicta. I disagree with Judge Alsup's suggestion on pirated books and Judge Chhabria's suggestion on copyright dilution, as more fully elaborated in my written statement. At this juncture, I think the best approach is for Congress to wait and see how other district courts, the courts of appeals, and potentially the Supreme Court resolve these difficult issues in the many pending copyright lawsuits.

---

[6] *Id.*; *see id.* at *10 ("First, an LLM's consumption of a book is different than a person's. An LLM ingests text to learn 'statistical patterns' of how words are used together in different contexts. It does so by taking a piece of text from its training data, removing a word from that text, predicting what that word will be, and updating its general understanding of language based on whether it was right or wrong—and then repeating this exercise billions or trillions of times with different text. This is not how a human reads a book. Second, unlike the hypothetical professor, Meta did not just give the plaintiffs' books to one person. Meta copied the plaintiffs' books as part of an effort to create a tool that can generate a wide range of text.").

[7] *Bartz*, 2025 WL 1741691, at *18.

[8] Lee, *Origin of AI Training*, at 149, 152, 156, 170-76, & nn. 229-51, 326-3 (tracing history of AI research and discovery of scaling by researchers, including citations of AI research articles).

[9] Bartz, 2025 WL 1741691, at *7; *Kadrey*, 2025 WL 1752484, at *15 ("Llama does not allow users to generate any meaningful portion of the plaintiffs' books. Neither party's expert opined that Llama was able to regurgitate more than 50 words from any of the plaintiffs' books, even in response to 'adversarial' prompting designed specifically to make LLMs regurgitate.").

[10] Executive Order, *Removing Barriers to American Leadership in Artificial Intelligence*, WHITE HOUSE (Jan. 23, 2025).

[11] New Generation Artificial Intelligence Development Plan (2017) (issued by State Council on Jul. 20, 2017).

[12] Google LLC v. Oracle Am., Inc., 593 U.S. 1, 20 (2021).

EXECUTIVE SUMMARY

- **Highly transformative purpose of AI training.** Judges Alsup and Chhabria correctly concluded, in *Bartz v. Anthropic* and *Kadrey v. Meta*, respectively, that the use of copies to train an AI model serves a *highly transformative* purpose in developing a technology under Factor 1 of fair use.

- **Origin of scaling by university researchers:** The history of university researchers training AI models on larger and more diverse datasets—a process called *scaling*, which proved to be a seminal breakthrough that led to the advances in AI today—supports this finding of a transformative purpose.

- **U.S. national interest in AI development:** President Trump's Executive Order declaring AI development a U.S. national priority and the Supreme Court's precedents recognizing that fair use is fact-specific and that copyright law must balance copyright and innovation both counsel caution and the avoidance of overbroad rulings or amendments that might jeopardize the U.S. national interest in AI development.

- **Need for caution:** Accordingly, I disagree with Judge Alsup's suggestion, in dicta, that pirated books are "irredeemably" infringing no matter the transformative purpose. And I disagree with Judge Chhabria's suggestion that most AI training is illegal under a new theory of market dilution. Neither categorical approach finds support in the text of the Copyright Act or case law.

- **No legislation needed at this time:** At this early stage of the copyright litigation involving AI companies, the best course for Congress is to wait and see how the cases are resolved by other district courts, the courts of appeals, and potentially the Supreme Court. *Bartz* and *Kadrey* are just two of more than forty AI copyright lawsuits.

I. THE ORIGIN OF AI TRAINING AND ITS TRANSFORMATIVE PURPOSE

More than forty copyright lawsuits against AI companies are now pending in the United States.[13] A central question in these copyright lawsuits is whether an AI company's use of copyrighted works to train and develop its AI models is a fair use—or not. In two different cases, federal judges, Judge Alsup in *Bartz v. Anthropic* and Judge Chhabria in *Kadrey v. Meta*, recently held that an AI company's use of copyrighted works to train AI models served a highly transformative purpose in developing the AI technology and, after balancing the four factors of fair use, the use was a fair use.[14] (Both judges were critical of the AI companies in non-precedential parts of their opinions related to acquiring pirated books and market dilution, respectively. I discuss these issues in Parts II and III.)

I believe Judges Alsup and Chhabria correctly concluded that the use of copyrighted works to train AI models serves a highly transformative purpose under Factor 1 of fair use, which

[13] *See Master List of Lawsuits v. AI*, CHATGPT IS EATING THE WORLD (updated Jun. 30, 2025).
[14] *See Bartz*, 2025 WL 1741691, at *7, *18; *Kadrey*, 2025 WL 1752484, at *9, *23.

examines the purpose and character of the defendant's use of copyrighted works.[15] Here, the purpose was to train an AI model and develop an innovative new technology. This "further purpose" in AI training is different from the authors' purpose in creating their books "for entertainment or education."[16] During training, the AI model is exposed to vast training materials, typically many millions of works, to identify the "statistical relationships among words." From this deep learning, the model develops the ability to conduct numerous functions, including research, translation of foreign languages, delivery of medical advice, generation of content, and so forth.[17] Critically, the plaintiffs in both cases did *not* show that the AI models had produced any infringing outputs of their books or any substantially similar copies.[18] Treating as fair use the creation of a new technology that does not redistribute significant portions of any works used in its development is amply supported by past fair use decisions, including *Google v. Oracle* and *Authors Guild v. Google*, as summarized in Appendices B and C and distinguished from cases in Appendix D involving technologies that merely redistributed infringing copies.[19]

As Judge Chhabria concluded, "The purpose of Meta's copying was to train its LLMs, which are innovative tools that can be used to generate diverse text and perform a wide range of functions."[20] And, as Judge Alsup recognized, "The technology at issue was among the most transformative many of us will see in our lifetimes."[21]

The history of AI development strongly supports this conclusion. It is important to understand *why* AI researchers at universities began training AI models on large datasets, much of which contained numerous copyrighted works used without permission. The practice originated, not at AI companies, but at universities where AI researchers discovered a key insight: *scaling*, or using larger and more diverse datasets actually worked in developing and improving AI models.[22] This seminal breakthrough, which took decades to figure out, propelled the advances in AI witnessed today.

Some uses might not be fair, however. A highly transformative purpose in AI training does not guarantee such use is a fair use, however. I agree with Judge Alsup's and Judge Chhabria's respective findings of fair use in the particular facts of the cases *Bartz* and *Kadrey*. But these decisions do not mean that AI training is fair use in every case. Courts must balance all four

---

[15] 17 U.S.C. § 107. *See Bartz*, 2025 WL 1741691, at *7; *Kadrey*, 2025 WL 1752484, at *9.
[16] *Kadrey*, 2025 WL 1741691, at *9-*10.
[17] *See id.* at *5, *10.
[18] *See Bartz*, 2025 WL 1741691, at *7; *Kadrey*, 2025 WL 1752484, at *9, *15.
[19] *See* Google LLC v. Oracle Am., Inc., 593 U.S. 1, 33 (2021) ("Here Google's use of the Sun Java API seeks to create new products. It seeks to expand the use and usefulness of Android-based smartphones. Its new product offers programmers a highly creative and innovative tool for a smartphone environment. To the extent that Google used parts of the Sun Java API to create a new platform that could be readily used by programmers, its use was consistent with that creative 'progress' that is the basic constitutional objective of copyright itself."); Authors Guild v. Google, Inc., 804 F.3d 202, 216, 224 (2d Cir. 2015) ("Google's making of a digital copy of Plaintiffs' books for the purpose of enabling a search for identification of books containing a term of interest to the searcher involves a highly transformative purpose, in the sense intended by *Campbell.*"; Google Book Search's snippet copy of small parts of the books in search results did not produce "meaningful or significant effect" of cognizable market harm, even though it could result in "some loss of sales" of books due to a user's ability to find an unprotected historical fact contained in the book).
[20] *Kadrey*, 2025 WL 1741691, at *9.
[21] *Bartz*, 2025 WL 1741691, at *18.
[22] Lee, *Origin of AI Training*, at 149, 152, 156, 170-76, & nn. 229-51, 326-3 (tracing history of AI research and discovery of scaling by researchers, including citations of AI research articles).

factors of fair use under the facts of each case, including the potential market harm to the copyright holder's original work and derivative works.

As explained in my scholarship, an AI model that routinely produces outputs that are infringing, such as regurgitations, might not be a fair use due to insufficient guardrails (and using more of the works than reasonably necessary for the purpose of developing an AI model).[23] Moreover, the outputs of an AI model are separate uses and can constitute separate acts of infringement if they are substantially similar copies of works in the training datasets.[24] And, in some cases, AI outputs that copy a specific artist's style—generated "in the style of" an artist—may well include copyrightable elements to support an infringement claim.[25]

II. HOW TO WEIGH THE USE OF PIRATED BOOKS FROM SHADOW LIBRARIES

My second recommendation relates to the controversial issue related to the use of pirated books from shadow libraries online. These shadow libraries were created by unnamed people who have escaped legal efforts to shut the sites down, even in the face of court orders.[26] Even when a website is blocked, a shadow library can easily resurface at a different site hosted by foreign locations.[27] The most contentious issue in the book author lawsuits is whether the AI company's acquiring and copying pirated books from shadow libraries online was (i) a separate infringing use or (ii) a use for the further purpose to train the defendant's AI models.[28]

Judges Alsup and Chhabria disagreed on this issue. Judge Alsup treated Anthropic's acquisition of copies from shadow library as a separate use for Anthropic's own library building—and held that such library building was not fair use.[29] In dicta, Judge Alsup suggested an even more categorical approach—that *any* acquiring of pirated copies was "inherently, irredeemably infringing" no matter what the transformative purpose and even if the copies were "immediately discarded" after a transformative use.[30] By contrast, Judge Chhabria took a flexible approach viewing the acquisition of the copies in relation to the defendant's further, transformative purpose in acquiring them, namely, AI training.[31] But Judge Chhabria recognized that the use of

---

[23] *Id.* at 213-15.
[24] *Id.* at 113 ("Under *Warhol*, courts can find that uses in AI training serve a fair purpose, but uses in AI outputs that are 'regurgitated' or substantially similar copies do not.").
[25] *See id.* at 202 ("Granted, some AI generators may copy copyrightable elements when generating a work in response to a person's prompt to create in 'the style of' a specific artist. But the proper remedy is a copyright infringement lawsuit, not concocting a mutant species of copyright dilution that penalizes non-infringing works."); *see also* 2 PATRY ON COPYRIGHT § 4:14 (2025) (distinguishing between unprotectable communal or generalized styles versus a style distinctive to an individual based on copyrightable elements of specific works).
[26] *See* Ashley Belanger, *"Most notorious" illegal shadow library sued by textbook publishers*, ARS TECHNICA (Sep. 15, 2023).
[27] *Id.*
[28] AI researchers determined that using books provided high-quality data to train LLMs that yielded better LLMs. For example, in internal emails disclosed as part of Kadrey's summary judgment motion, Meta developers concluded that the use of the controversial Library Genesis dataset is "essential to meet SOTA [state of the art] numbers across all categories." *Evidence of Meta's use of LibGen dataset and seeding torrents to share files. Wanted to compete with OpenAI and Mistral*., CHATGPT IS EATING THE WORLD (Feb. 6, 2025); *see* Kadrey v. Meta Platforms, Inc., -- F. Supp. 3d --, 2025 WL 1752484, at *5 (N.D. Cal. June 25, 2025) ("To be able to generate a wide range of text—in different languages or styles, or regarding different subject matter—an LLM's training dataset must be large and diverse…. But while a variety of text is necessary for training, books make for especially valuable training data. This is because they provide very high-quality data for training an LLM's 'memory' and allowing it to work with larger amounts of text at once.").
[29] *See* Bartz v. Anthropic PBC, -- F. Supp. 3d --, 2025 WL 1741691, at *11-*14 (N.D. Cal. Jun. 23, 2025).
[30] *Id.* at *11.
[31] *Kadrey*, 2025 WL 1752484, at *12.

pirated books can weigh *against* fair use as market harm if the evidence showed the "use of shadow libraries benefited those libraries or their other users."[32] In *Kadrey*, Judge Chhabria concluded the plaintiffs failed to present sufficient evidence of such market harm.[33]

Judge Chhabria's flexible approach to pirated copies offers the better way for courts to address the issue of pirated books. First, it is faithful to the text of the Copyright Act. Unlike the first-sale doctrine and other copyright exceptions, the text of Section 107, the fair use provision, contains no requirement that the defendant use a "lawfully made copy" to qualify for fair use.[34] In enacting the Copyright Act of 1976, Congress knew how to draft a per se requirement of a "lawfully made copy" for a copyright exception, but did not do so for fair use.[35] The text of Section 107 forecloses the adoption of any per se requirement that people must obtain a lawfully made copy to assert a valid fair use defense.

Second, when it had a chance to recognize such a per se requirement in *Harper & Row*, which involved a purloined manuscript of a book, the Supreme Court did not do so—instead weighing the purloined character in the overall balance of fair use factors.[36] In *Google v. Oracle*, the Court also declined to recognize "bad faith" of the defendant as a relevant factor, while quoting from Judge Leval's seminal fair use article that "[c]opyright is not a privilege reserved for the well-behaved."[37] The *Warhol* Court recognized that "[m]ost copying has some further purpose."[38] A defendant should not be precluded from asserting a further purpose to justify making an unauthorized copy as a part of a fair use defense, but the defendant faces potential liability if the defense fails. Likewise, the Copyright Office's pre-publication report on AI training rejected a categorical approach but instead recommended that the use of pirated copies "should weigh against fair use without being determinative."[39] Third, Judge Chhabria's approach to pirated books leaves open the possibility that an AI company's use of them will *not* be fair use based on the submission of evidence that such use materially supported a shadow library.[40] Far from

---

[32] *Id.*
[33] *See id.* at *21. ("But although the plaintiffs discussed Meta's use of shadow libraries at length, they did not argue that it had these effects or was relevant to the fourth factor beyond allowing Meta to get the books without paying….[T]he plaintiffs' counsel did suggest that, by using shadow libraries, Meta (and other companies like it) would reduce the stigma associated with shadow libraries and encourage more people to use them. It's not clear whether this would matter in the overall analysis. But in any event, counsel conceded that the record contains no evidence of this dynamic playing out.") (internal citation omitted).
[34] *Compare* 17 U.S.C. § 109(a) ("the owner of a particular *copy … lawfully made* under this title") (emphasis added) *with id.* § 107 ("fair use of a copyrighted work"); Kirtsaeng v. John Wiley & Sons, Inc., 568 U.S. 519, 537 (2013) (discussing "lawfully made" copy requirement in §§ 109(c) (exception to public display), 109(e) (exception for video games in coin-operated equipment), and 110(1) (in-classroom teaching exception to public display and performance but not if copy "not lawfully made"); *see also* 17 U.S.C. § 108(c)(2) ("lawful possession of such copy" by library or archives).
[35] *Kirtsaeng*, 568 U.S. at 537 (the prior 1909 Copyright Act's language for the first-sale doctrine was even more explicit in requiring for the first-sale exception: "'[N]othing in this Act shall be deemed to forbid, prevent, or restrict the transfer of any copy of a copyrighted work *the possession of which has been lawfully obtained.*'" Copyright Act of 1909, § 41, 35 Stat. 1084 (emphasis added)).")
[36] Harper & Row Publishers, Inc. v. Nation Enters., 471 U.S. 539, 563 (1985).
[37] Google LLC v. Oracle Am., Inc., 593 U.S. 1, 32-33 (2021) (quoting Pierre N. Leval, *Toward a Fair Use Standard*, 103 HARV. L. REV. 1105, 1126 (1990)).
[38] Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith, 598 U.S. 508, 528-29 (2023).
[39] U.S. Copyright Office, *Copyright and Artificial Intelligence Part 3: Generative AI Training* (Pre-Publication Report) 52 (May 2025).
[40] *Kadrey*, 2025 WL 1752484, at *12, *21.

giving a green light to the use of pirated books datasets, Judge Chhabria's approach signals great legal risk for any AI company that does so.[41]

This fact-specific approach allows courts to carefully balance the four factors of fair use. Granted, the file sharing of unauthorized copies of works is infringement in many cases. But Section 107 and Supreme Court precedent do not support a rigid, categorical approach to treat every unauthorized copy as "inherently, irredeemably infringing" no matter what the transformative purpose the defendant had. Even copying for the purpose of library-building, the Supreme Court in *Grokster* said was not "necessarily infringing."[42] Even in the illegal music-file sharing cases, the courts initially evaluated the defendant's asserted purpose, such as the practice of sampling to decide whether to purchase a copy or the convenience of space-shifting in digital format, but ultimately held it was not transformative.[43] By contrast, in both *Bartz* and *Kadrey*, both judges found the defendant's use in AI training was highly transformative.[44] And the notion that property that was initially illicit is "irredeemably" so and can never be repurposed for legitimate public ends is not recognized in other areas of federal law.[45]

Assume for the sake of argument the courts in a decision or Congress in an amendment adopts a per se requirement that a defendant must initially acquire a "lawfully made copy" of a work to be able to assert a fair use defense. This categorical approach would dramatically shrink the scope of fair use. *Every* fair use necessarily involves an unauthorized copy—indeed, that is very question whether the unauthorized copy is a fair use. Judge Alsup's opinion repeatedly referred to "pirated book" without explaining the term, much less whether it is different from an

---

[41] Moreover, evolving norms in AI training might coalesce around some best practices. *See The EU Code of Practice and future of AI in Europe*, OPENAI (Jul. 11, 2025) (intent to sign EU's Code of Practice for General Purpose of AI); EU Code of Practice for General-Purpose AI Models, Copyright Chapter, Measure 1.2 (measure to copy "only lawfully accessible copyright-protected content when crawling the World Wide Web").

[42] MGM Studios, Inc. v. Grokster, Ltd., 545 U.S. 913, 931 (2005) (discussing video library building in Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417, 424, 454-55 (1984) as not necessarily infringing).

[43] *See, e.g.*, BMG Music v. Gonzalez, 430 F.3d 888, 890 (7th Cir. 2005) (sampling purpose failed because "[i]nstead of erasing songs that she decided not to buy, she retained them."); A&M Records, Inc. Napster, Inc., 239 F.3d 1004, 1018-19 (9th Cir. 2001) (space shifting purpose failed because file-sharing also involved distributing music files to others); *see also* UMG Recordings, Inc. v. MP3.com, 92 F. Supp. 2d 349, 351 (S.D.N.Y. 2000) (space shifting service was "simply another way of saying that the unauthorized copies are being retransmitted in another medium—an insufficient basis for any legitimate claim of transformation"). Based on these and other music-file sharing decisions, any defendant engaged in music file sharing of copyrighted works is likely engaging in infringement, with no fair use defense. *See In re DMCA § 512(h) Subpoena to Twitter, Inc.*, 608 F. Supp. 3d 868, 879 (N.D. Cal. 2022) (Chhabria, J.) ("In some cases, no analysis is required; it is obvious, for example, that downloading and distributing copyrighted music via peer-to-peer systems does not constitute fair use."); *see also* U.S. v. Slater, 348 F.3d 666, 668-69 (7th Cir. 2003) (upholding denial of jury instruction of fair use in criminal case against defendant who was participant in website to distribute illegally pirated software). These cases involve mere redistribution of copies of works, which, as shown in the table in Appendix D attached to this statement, is typically not fair use.

[44] Bartz v. Anthropic PBC, -- F. Supp. 3d --, 2025 WL 1741691, at *7 (N.D. Cal. Jun. 23, 2025); Kadrey v. Meta Platforms, Inc., -- F. Supp. 3d --, 2025 WL 1752484, at *9 (N.D. Cal. June 25, 2025).

[45] *See* Kristina Rae Montanaro, Note, *"Shelter Chic": Can the U.S. Government Make It Work*, 42 VAND. J. TRANSNAT'L L. 1663, 1664-65 (2009) (discussing U.S. Customs Service's donation of seized counterfeit goods for humanitarian relief after Hurricane Katrina, exceeding $20 million in value); *The US Government Sold Nearly 10,000 Silk Road Bitcoin*, NASDAQ (Mar. 31, 2023); *NCDA & Partners Mark End of Project Donating Nearly 100K Seized Counterfeit Jackets to NY Charities*, NASSAU CO. DISTRICT ATT'Y (Apr. 28, 2022) ("Nassau County District Attorney Anne T. Donnelly today announced the completion of a six-year long effort to donate nearly 100,000 counterfeit jackets – seized during multiple investigations – to more than 160 charities across Long Island and the greater New York area."); *Real Property Auctions*, U.S. TREASURY (listing auctions of real property seized by federal government); *Disposition of Seized, Forfeited, Voluntarily Abandoned, and Unclaimed Personal Property*, U.S. DEP'T OF INTERIOR (allowing donation of seized drug paraphernalia for law enforcement or educational purposes).

unauthorized initial copy.[46] If it is the same, then every dataset of copyrighted works collected for AI training is pirated—and every AI company's and every researcher's acquisition of the unauthorized dataset is infringement. Such an extreme result would greatly hinder AI development in the United States.

Even limited to the pirated books datasets online, Judge Alsup's suggested categorical approach to pirated books disproportionately favors Big Tech and other well-financed companies that have the resources to spend many millions of dollars to buy physical books and manually scan digital copies of them for AI training as Anthropic eventually did.[47] And, among Big Tech, Google might have a big advantage given its Google Book search database. Small tech companies and independent researchers would have little chance in contributing to innovation in AI models. Such a rule favoring Big Tech companies is bad for innovation in the United States.[48]

III. THE U.S. NATIONAL INTEREST IN AI DEVELOPMENT AND INNOVATION

The U.S. national interest in AI development and innovation counsels caution by the courts, Congress, and the states. Technological progress is just as important to the United States as artistic progress.[49] In three technology-related copyright cases, the Supreme Court recognized the important need to balance the competing interests in copyright and technological innovation.[50] As the Federal Circuit explained citing the legislative history of the Copyright Act of 1976, the fair use doctrine provides a way for courts to address "technological innovations."[51] The fair use doctrine is an American doctrine, which originated in the United States and has accommodated innovation from the VCR to programs that enhance the Internet and smartphones, all technologies of great national significance. In *Grokster,* the Court also described the *Sony* safe harbor—for technologies capable of substantial non-infringing uses—as a doctrine that "leaves breathing room for innovation and a vigorous commerce."[52]

In an executive order, President Trump declared: "It is the policy of the United States to sustain and enhance America's global AI dominance . . . to promote human flourishing, economic competitiveness, and national security."[53] Then-President Biden had recognized the same priority in AI in an earlier executive order.[54] And, in 2018, Congress established the independent National Security Commission on Artificial Intelligence, which warned in 2021: "For the first time since World War II, America's technological predominance—the backbone of its economic

---

[46] Edward Lee, *Judge Alsup's Solomonic judgment on fair use in AI training & acquiring pirated books: is it the blueprint for the future of AI training? Part I: Pirated copies*, CHATGPT IS EATING THE WORLD (June 25, 2025).
[47] *See Bartz,* 2025 WL 1741691, at *2.
[48] *See generally* Mark Lemley & Watt Wansley, *How Big Tech Is Killing Innovation*, N.Y. TIMES (Jun. 13, 2024); Mark A. Lemley & Matthew T. Wansley, *Coopting Disruption*, 105 BOSTON UNIV. L. REV. 458 (2025).
[49] *See* Lee, *Origin of AI Training*, at 145-47.
[50] *See* Metro-Goldwyn-Mayer Studios, Inc. v. Grokster, Ltd., 545 U.S. 913, 933 (2005); Google LLC v. Oracle Am., Inc., 593 U.S. 1, 22 (2021); Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417, 442-43 (1984); Lee, *Origin of AI Training*, at 126-29, 143-47.
[51] Atari Games Corp. v. Nintendo of Am., Inc., 975 F.2d 832, 843 (Fed. Cir. 1992); H.R. REP. NO. 94-1476, at 66 (1976) ("The bill endorses the purpose and general scope of the judicial doctrine of fair use, but there is no disposition to freeze the doctrine in the statute, especially during a period of rapid technological change."); S. REP. NO. 94-473, at 62 (1975) (same).
[52] *Grokster*, 545 U.S. at 933.
[53] President Donald J. Trump, Executive Order, *Removing Barriers to American Leadership in Artificial Intelligence*, WHITE HOUSE (Jan. 23, 2025).
[54] *FACT SHEET: Biden-Harris Administration Announces New AI Actions and Receives Additional Major Voluntary Commitment on AI*, Internet Archive (Jul. 26, 2024).

and military power—is under threat. China possesses the might, talent, and ambition to surpass the United States as the world's leader in AI in the next decade if current trends do not change."[55]

Although courts decide private disputes between parties, the consideration of fair use allows courts to consider the *public benefit* of the defendant's use, as well as how it may serve the overall constitutional goal of "promoting Progress" in the United States.[56] As the Supreme Court explained in another technology case, *Google v. Oracle*, the transformative purpose "to create a new platform" that enables others to create new applications was "consistent with that creative 'progress' that is the basic constitutional objective of copyright itself."[57] Indeed, Judges Alsup and Chhabria cited the Copyright Clause or the *Google* Court's analysis of the transformative purpose in developing a new computing platform.[58]

Given the U.S. national priority in AI innovation, both the courts and Congress should proceed cautiously before adopting a categorical or inflexible rule that might greatly hamper AI innovation. For example, in an extensive section of dicta, Judge Chhabria concluded that, in "most cases," AI training on copyrighted works is likely "generally illegal" and that AI "companies, to avoid liability for copyright infringement, will generally need to pay copyright holders for the right to use their materials."[59] To reach that sweeping conclusion, Judge Chhabria speculated for many pages on a new theory of copyright market dilution, even though he held that the plaintiffs had failed to present sufficient evidence of their own putative market harm to survive summary judgment.[60]

The new theory of copyright market dilution should be rejected. It impermissibly expands the scope of copyright to non-infringing works of others and treats those non-infringing works as cognizable market harm that a copyright holder can claim under fair use.[61] Thus, in Judge Chhabria's view, even if people using AI do not produce any infringing outputs, the fair use defense in training the respective AI model would still fail *simply because the model can produce non-infringing outputs in the same genre or type of work* in the training datasets.[62] For example, if an AI model was trained on romance novels, every non-infringing romance novel someone creates using that model can constitute market dilution under Factor 4—even though the non-infringing romance novel contains no copied protected expression from the works in the training datasets.[63] Judge Chhabria conceded that no court has ever recognized such an expansive view of market harm to include non-infringing works of others.[64]

---

[55] National Security Comm'n on Artificial Intelligence, *Final Report* 7 (2021).
[56] Google LLC v. Oracle Am., Inc., 593 U.S. 1, 35-36 (2021) ("Further, we must take into account the public benefits the copying will likely produce.").
[57] *Id.* at 30.
[58] *See* Bartz v. Anthropic PBC, -- F. Supp. 3d --, 2025 WL 1741691, at *8 (N.D. Cal. Jun. 23, 2025) (citing Copyright Clause, U.S. CONST. art I, § 8, cl. 8); Kadrey v. Meta Platforms, Inc., -- F. Supp. 3d --, 2025 WL 1752484, at *9 (N.D. Cal. June 25, 2025) (quoting and citing *Oracle*, 593 U.S. at 30).
[59] *Kadrey*, 2025 WL 1752484, at *1-*2.
[60] *Id.* at 1-*2, *15-*23.
[61] *Id.*
[62] *Id.* at *15.
[63] *Id.* at *18.
[64] *Id.* ("it's never made a difference in a case before").

As I explain at greater length in my law review article, copyright market dilution is overbroad and likely unconstitutional.[65] Dilution is a trademark concept. It did not become recognized under federal trademark law until Congress amended the Lanham Act in 1995, expanding the scope of trademarks for famous marks to prohibit dilution.[66] To apply dilution to expand the scope of copyrights in the fair use analysis is problematic. To borrow Justice Scalia's apt phrase in an analogous case, the new theory of copyright dilution is "a species of mutant copyright law": it misuses a trademark concept to protect copyrighted works.[67]

Market dilution extends copyright beyond the constitutional limitations in the Copyright Clause, which gives Congress the power to grant exclusive rights to authors only in "*their respective writings*."[68] A genre, type, or kind of work is not a "writing"; it is just an idea or method for identifying a stock or common way to organize expression, such as an article, essay, novel, poem, or play.[69] As Judge Alsup recognized in rejecting market dilution in *Bartz*, "Copyright does not extend to 'method[s] of operation, concept[s], [or] principle[s]' 'illustrated[ ] or embodied in [a] work.'"[70] Indeed, as Justice Story explained, "Every book in literature, science and art, borrows, and *must necessarily borrow*, and use much which was well known and used before."[71] This fundamental limitation on copyright explains why no novelist can own the genre for novels, romance or otherwise, let alone claim cognizable market harm from other novels that do not infringe. Yet the theory of mutant copyright dilution proposes to do just that—expand copyright to genres and uncopyrightable ideas.[72] Such an expansion turns the Copyright Clause on its head. Instead of promoting progress, the goal is to protect copyrights—and perversely to reduce the creation of new, non-infringing works.

---

[65] See Lee, *Origin of AI Training*, at 188-208.
[66] Erin J. Roth & Robert B. Bennett, Jr., *The Federal Trademark Dilution Act: Potent Weapon or Uphill Battle*, 16 MIDWEST L. REV. 1, 7-11 (1999) (history of Congress's enactment of dilution protection for famous marks in Lanham Act in 1995).
[67] Dastar Corp. v. Twentieth Century Fox Film Corp., 539 U.S. 23, 34 (2003).
[68] U.S. CONST. art. I, § 8, cl. 8.
[69] *See* Hassett v. Hasselbeck, 757 F. Supp. 2d 73, 89-90 (D. Mass. 2010) ("While the books address some of the same topics, the order of presentation is not identical or nearly so. To the extent there is any general similarity related to the selection and ordering of the topics, the defendants' exhibits demonstrate that the general sequence and topic selection of these works are customary to the genre, and thus unprotected under the doctrine of *scenes a faire. See Coquico,* 562 F.3d at 68. Moreover, courts have held that the general thematic ordering and arrangement of a work is not usually copyrightable. *See LaPine,* 2009 WL 2902584, at *9; *see also Dunn,* 517 F.Supp.2d at 544 (holding that the claim that two works have substantial thematic and structural similarity 'has little or no support in the law as a basis for a copyright claim'). To the extent there is any similarity between the structures of the two works, that similarity relates to unprotected elements of the works and does not support a finding of substantial similarity."); *see also* Nichols v. Universal Pictures Corp., 45 F.2d 119, 121-22 (2d Cir. 1930) (discussing how copyright does not protect unprotected ideas and elements in a play); Peters v. West, 692 F.3d 629, 636 (7th Cir. 2012) ("Copyright protects actual expression, not methods of expression. 17 U.S.C. § 102(b); Baker v. Selden, 101 U.S. 99 (1879). Just as a photographer cannot claim copyright in the use of a particular aperture and exposure setting on a given lens, no poet can claim copyright protection in the form of a sonnet or a limerick.").
[70] Bartz v. Anthropic PBC, -- F. Supp. 3d --, 2025 WL 1741691, at *8 (N.D. Cal. Jun. 23, 2025) (quoting 17 U.S.C. § 102(b)).
[71] Emerson v. Davies, 8 F. Cas. 615, 619 (No. 4,436 (CCD Mass. 1845) (Story, J.) (emphasis added); *see* Zechariah Chafee, *Reflections on the Law of Copyright*, 45 COLUM. L. REV. 503, 511 (1945) ("Progress would be stifled if the author had a complete monopoly of everything in his book for fifty-six years or any other long period. Some use of its contents must be permitted in connection with the independent creation of other authors. The very policy which leads the law to encourage his creativeness also justifies it in facilitating the creativeness of others.").
[72] *Cf.* Design Basics, LLC v. Signature Construction, Inc., 994 F.3d 879, 889 (7th Cir. 2021) ("Standard elements in a genre—called *scènes à faire* in copyright law—get no copyright protection. *Scènes à faire* are 'so rudimentary, commonplace, standard, or unavoidable that they do not serve to distinguish one work within a class of works from another.'" Bucklew v. Hawkins, Ash, Baptie & Co., 329 F.3d 923, 929 (7th Cir. 2003). If standard elements received copyright protection, then the creation of a single work in a genre would prevent others from contributing to that genre because the copyright owner would have exclusive rights in all of the genre's basic elements.").

That result also violates the First Amendment under which "*more speech*, not less, is the governing rule."[73] Copyright market dilution seeks to protect copyrights and *reduce* new expression embodied in non-infringing works people created using AI.[74] It penalizes, under fair use, *non-infringing expression* of others—with the likely consequence of making illegal the very AI technology that people used to create the non-infringing expression. As Judge Chhabria openly stated, most AI training will be *generally illegal* under market dilution.[75] If so, then the First Amendment rights of many people who use AI will be impaired. Such a radical change to the traditional contours of copyright, fair use, and the idea-expression dichotomy requires strict scrutiny under the First Amendment.[76]

Courts can no more treat as dilution the *non-infringing* expression of others under copyright law simply because they used AI tools than courts can treat as defamation people's *truthful* expression simply because they used AI tools in creating the expression. The First Amendment protects non-infringing expression and truthful expression alike.[77]

In applying fair use, courts must balance competing interests, including the larger public interest and benefits.[78] Heeding the Supreme Court's fair use precedents counsels caution. As the *Google* Court admonished in another technology case, "Given the rapidly changing technological, economic, and business-related circumstances, we believe we should not answer more than is necessary to resolve the parties' dispute."[79]

The extensive dicta in *Kadrey* on a new, expansive theory of copyright dilution based solely on non-infringing expression—which is protected by the First Amendment—failed to follow the Supreme Court's admonition. Instead, it opined that, in most cases, AI training is outright illegal. Such a radical categorical approach flouts the Supreme Court's repeated avoidance of bright-line rules in applying fair use.[80] And, if adopted, it jeopardizes the United States' national interest in AI. As the White House AI Czar David Sacks advised, "There must be a fair use concept for training data or models would be crippled. China is going to train on all the data regardless, so

---

[73] Citizens United v. FEC, 558 U.S. 310, 361 (2010) (emphasis added).
[74] See Lee, *Origin of AI Training*, at 202, 204-05.
[75] *See Kadrey*, 2025 WL 1752484, at *1; *id.* at *2 ("The upshot is that in many circumstances it will be illegal to copy copyright-protected works to train generative AI models without permission. Which means that the companies, to avoid liability for copyright infringement, will generally need to pay copyright holders for the right to use their materials.")
[76] *See* Eldred v. Ashcroft, 537 U.S. 186, 191 (2003) ("When, as in this case, Congress has not altered the traditional contours of copyright protection, further First Amendment scrutiny is unnecessary."); Lee, *Origin of AI Training*, at 208-09.
[77] *See* Pan Am Sys., Inc. v. Atlantic Northeast Rails & Ports, Inc., 804 F.3d 59, 65-66 (1st Cir. 2015) (First Amendment protects truthful information, which is complete defense to defamation); Neil Weinstock Netanel, *First Amendment Constraints on Copyright after* Golan v. Holder, 60 UCLA L. Rev. 1082, 1128 2013) ("Courts must, accordingly, interpret and apply the idea/expression dichotomy and fair use privilege in a manner consistent with their vital First Amendment role. Further, following *Golan*, statutory provisions that disturb copyright's built-in First Amendment accommodations, or that otherwise abridge noninfringing speech, lie vulnerable to First Amendment challenge.").
[78] *See* Google LLC v. Oracle Am., Inc., 593 U.S. 1, 35-36 (2021); *see also* Metro-Goldwyn-Mayer Studios, Inc. v. Grokster, Ltd., 545 U.S. 913, 928 (2005) ("The more artistic protection is favored, the more technological innovation may be discouraged; the administration of copyright law is an exercise in managing the tradeoff."); Goldstein v. California, 412 U.S. 546, 559 (1973) ("Where the need for free and unrestricted distribution of a writing is thought to be required by the national interest, the Copyright Clause and the Commerce Clause would allow Congress to eschew all protection.").
[79] *Google*, 593 U.S. at 21.
[80] *See* Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith, 598 U.S. 508, 528 (2023); Google LLC v. Oracle Am., Inc., 593 U.S. 1, 18-19 (2021); Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 577 (1994).

without fair use, the U.S. would lose the AI race."[81] Indeed, already one court in China, stressing the need for "*encourag[ing] technological progress*," indicated that the use of copyrighted works to train an AI model is permissible under Chinese copyright law provided the AI model does not produce infringing outputs, in a decision upheld on appeal.[82]

CONCLUSION

Any categorical approach that would make all AI training illegal and not fair use—such as the dicta in Judge Chhabria's opinion in *Kadrey v. Meta* concluding that most AI training is illegal under the speculative new theory of copyright market dilution—should be rejected. Such a ruling is contrary to the Copyright Clause's limitation of copyright to only authors' "respective writings," the Copyright Act's exclusion of ideas, methods, and genres from protection, and the Supreme Court's repeated admonition on the fact-specific nature of fair use. If adopted, such a ruling would not only hamper AI innovation in the United States, but it also may prompt U.S. companies to relocate their AI training offshore to countries with copyright exceptions for text-data-mining (TDM) or fair use that would allow AI training, including Israel, Japan, and Singapore.[83] The fair use provision Congress codified in the Copyright Act as a flexible doctrine to accommodate "rapid technological change" does not support, much less require, such a drastic result.[84]

---

[81] @DavidSacks, X (Jun. 24, 2025, 10:10 AM), https://x.com/davidsacks/status/1937558998166954092; *see* National Security Comm'n on Artificial Intelligence, *Final Report* 2, 4 (2021) ("But we must win the AI competition that is intensifying strategic competition with China. China's plans, resources, and progress should concern all Americans. It is an AI peer in many areas and an AI leader in some applications. We take seriously China's ambition to surpass the United States as the world's AI leader within a decade….The federal government must partner with U.S. companies to preserve American leadership and to support development of diverse AI applications that advance the national interest in the broadest sense."). In 2021, China amended its copyright act to include, in clause 13 to Article 24, a general exception for "[o]ther circumstances provided by laws and administrative regulations." Matthew Sag & Peter K. Yu, *The Globalization of Copyright Exceptions for AI Training*, 74 EMORY L. REV. 1163, 1194 (2025). Although the provision has not yet been applied to this circumstance, some legal commentators suggest it could support AI training in China. *See id.*

[82] *See* Shanghai XX Cultural Dev. Co. v. Hangzhou XX Intelligent Tech. Co., (2024) Zhe 0192 Min Chu 1587 (Hangzhou Internet Ct. Sept. 25, 2024), a*ff'd*, (2024) Zhe 01 Mín Zhong 10332 (Hangzhou Interm. People's Ct. Dec. 30, 2024) ("Therefore, this court believes that It can be considered [permissible] use when there is no evidence that generative artificial intelligence is for the purpose of using the original expression of the right work, has affected the normal use of the right work, or unreasonably harms the legitimate interests of the relevant copyright holders.") (bracketed translation of "permissible use" inserted); *id.* ("Finally, there should be a prudent and inclusive approach to generative AI that *encourages technological progress and business development*. The creation and development of generative artificial intelligence requires the introduction of huge amounts of training data at the input end, which is unavoidable [to avoid] using other people's works. In view of the purpose of generative AI to use other people's works in the data training stage, it should in principle be used to learn and analyze the thoughts, feelings, language features, characteristic styles, etc. expressed in previous works, and extract corresponding rules, structures, patterns, and trends from them to facilitate subsequent transformational creation of new works. This kind of 'use behavior' to aggregate a large number of works as analysis sample data for training to improve the creative ability of the work is not for the purpose of reproducing the original expression of the work, and generally the data training is only to temporarily retain the previous work when analyzing the structural characteristics of the corpus data, the data training and generation process did not display the previous works to the public. Therefore, this court believes that it can be considered [permissible] use when there is no evidence that generative.") (bracketed translation of "permissible use" inserted and emphasis added). An AI platform was held secondarily liable for allowing infringing outputs of the character Ultraman. *See* King & Wood Mallesons, *Chinese AIGC Platform Found Secondarily Liable for Copyright Infringement*, LEXOLOGY (Feb. 28, 2025).

[83] *See* Sag and Yu, *supra*, at 1179-80, 1185-92; Jonathan Band, *Israel Ministry of Justice Issues Opinion Supporting Use of Copyrighted Works for Machine Learning*, DISRUPTIVE COMPETITION PROJECT (Jan. 19, 2023).

[84] *See* H.R. REP. NO. 94-1476, at 66 (1976).

APPENDIX A: Comparison of Decisions by Judge Alsup & Chhabria

FAIR USE FACTOR	JUDGE ALSUP Bartz v. Anthropic PBC	JUDGE CHHABRIA Kadrey v. Meta Platforms
Is downloading "pirated" books datasets separate use from training model?	***Separate use for library:*** Anthropic downloading / building a permanent library of pirated books was *not fair use*. Infringing. Trial on damages.	***Same use to train*:** Meta downloading was for the further purpose of training AI model. Fair use (but would not be had plaintiffs proven market dilution).
(1) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes	***Favors fair use (+).*** Train LLMs is **exceedingly** "**transformative — spectacularly so**" because it maps statistical relationships to produce technology that produces new, noninfringing outputs. This technology is "among the most transformative many of us will see in our lifetimes." No allegation of output infringing authors' works. Cites *Google Books* decision.	***Favors fair use (+).*** Meta's use of the plaintiffs' books had a **"further purpose" and "different character" than the books—that it was highly transformative.** "The purpose of Meta's copying was to train its LLMs, which are innovative tools that can be used to generate diverse text and perform a wide range of functions….The purpose of the plaintiffs' books, by contrast, is to be read for entertainment or education." Commercial use tends to be less important when the secondary use is highly transformative. Cites *Oracle* decision.
(2) the nature of the copyrighted work	***Disfavors fair use (-).*** Creative expressive books.	***Disfavors fair use (-).*** But second factor weighs less if transformative purpose.
(3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole	***Favors fair use (+).*** "Compelling benefits of training the LLMs on strong examples were not offset by revelations to the public of any portion of the works themselves. What was copied was therefore especially reasonable and compelling."	***Favors fair use (+).*** "The amount that Meta copied was reasonable given its relationship to Meta's transformative purpose. *See Oracle*, 593 U.S. at 34. Everyone agrees that LLMs work better if trained on more high-quality material."
(4) the effect of the use upon the potential market for or value of the copyrighted work.	***Favors fair use (+).*** No cognizable market harm. Authors concede that training LLMs did not result in any exact copies nor even infringing knockoffs of their works being provided to the public. ***Rejects new theory of *copyright market dilution*.** The Copyright Act seeks to advance original works of authorship, **not to protect authors against competition.** Rejects lost licensing as a market Copyright Act entitles authors to exploit.	***Favors fair use (+).**** *Slight public benefit.* Likely help Llama create new expression. ***But accepts new theory of *copyright market dilution*:** harm by helping to enable the rapid generation of countless works that compete with the originals, **even if those works aren't themselves infringing.** But finds Plaintiffs failed to present sufficient evidence to create genuine issue. Rejects lost licensing as a market authors entitled to exploit. Circularity problem.
JUDGE'S CONCLUSION	**FAIR USE** (but not for library)	**FAIR USE** (but not if market dilution)

APPENDIX B: Technology Fair Use Decisions Decided or Favorably Cited by Supreme Court

Edward Lee, *Fair Use and the Origin of AI Training*, 63 HOU. L. REV. (forthcoming 2025) (table 2).

CASES	TECHNOLOGY DEVELOPMENT: Use of Copyrighted Works to Create New Technology?	TECHNOLOGY USAGE: Use of Copyrighted Works in Public Use of Technology?	FACTOR 1: Use of Copyrighted Works Had Further Purpose or Different Character?
Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417 (1984).	No	Yes, by users of VCR for personal time-shift recordings.	No, time-shifted copies of free TV shows.
Google LLC v. Oracle Am., Inc., 593 U.S. 1 (2021).	Yes, use of Java declaring code for Android operating system to facilitate computer programmers' ability to write apps for Android.	Yes, declaring code was part of Android OS and can be used by programmers writing Android apps.	Yes, "use of the Sun Java API seeks to create new products," i.e., "a highly creative and innovative tool for a smartphone environment."
Sega Enters. Ltd. v. Accolade, Inc., 977 F.2d 1510 (9th Cir. 1992).	Yes, in reverse engineering of OS to find uncopyrightable element necessary for interoperability of new game.	No	Yes, "intermediate copying of computer code as an initial step *in the development of a competing product.*"
Sony Comput. Ent., Inc. v. Connectix Corp., 203 F.3d 596 (9th Cir. 2000).	Yes, in reverse engineering of OS to find uncopyrightable element necessary for game emulator to make games run on PC.	No	Yes, "*creates a new platform*, the personal computer, on which consumers can play games designed for the Sony PlayStation."
Authors Guild v. Google, Inc., 804 F.3d 202 (2d Cir. 2015).	Yes, to create a database to enable within-book search of published books and to enable text data mining analysis of frequency of use of words in entire corpus of books.	Yes, copies stored in database and snippets of books shown.	Yes, copies in database serve further purpose of searching within-text of all books in database to find relevant sources.

APPENDIX C: Lower Courts' Finding Fair Use in Other Technology Cases

Edward Lee, *Fair Use and the Origin of AI Training*, 63 HOU. L. REV. (forthcoming 2025) (table 3).

CASES	TECHNOLOGY DEVELOPMENT: Use of Copyrighted Works to Create New Technology?	TECHNOLOGY USAGE: Use of Copyrighted Works in Public Use of Technology?	FACTOR 1: Use of Copyrighted Works Had Further Purpose or Different Character?
Kelly v. Arriba Soft Corp., 336 F.3d 811, (9th Cir. 2003).	Yes, to create a searchable database of online images	Yes, copies stored in database and outputs show thumbnail images	Yes, copies in database serve further purpose of searching online images to find relevant ones.
Perfect 10, Inc. v. Amazon.com, Inc., 508 F.3d 1146 (9th Cir. 2007).	Yes, to create a searchable database of online images	Yes, copies stored in database and outputs show thumbnail images of reduced resolution	Yes, copies in database serve further purpose of searching online images to find relevant ones.
A.V. ex rel. Vanderhye v. iParadigms, LLC, 562 F.3d 630 (4th Cir. 2009).	Yes, to create a searchable database of student papers	Yes, copies stored in database but no direct quotations	Yes, copies in database serve further purpose of finding potential plagiarism in student papers
Authors Guild, Inc. v. HathiTrust, 755 F.3d 87, (2d Cir. 2014).	Yes, to create a database to enable within-book search of published books. Secondary use to store digital copies for preservation.	Yes, copies stored in database. No snippets of books shown, but (i) pages numbers in books term is found; (ii) access to full books to people with print-disability.	Yes, copies in database serve further purpose of searching within-text of all books in database to find relevant sources.
Field v. Google, Inc., 412 F. Supp. 2d 1106 (D. Nev. 2006).	Yes, to create a searchable database of cached copies of Internet websites	Yes, copies stored in database and "cached" static copy of website publicly accessible.	Yes, copies in database serve further purpose of allowing static view of snapshot of website, useful when website is down
White v. West Pub. Corp., 29 F. Supp. 3d 396 (S.D.N.Y. 2014).	Yes, to create a searchable database of copies of legal briefs to "creat[e] an interactive legal research tool."	Yes, copies stored in databases of West and Lexis.	Yes, copies in database serve further purpose of "creating interactive legal research tool."

APPENDIX D: Courts' Rejection of Fair Use in Technology Cases

Edward Lee, *Fair Use and the Origin of AI Training*, 63 HOU. L. REV. (forthcoming 2025) (table 4).

CASES	TECHNOLOGY DEVELOPMENT: Use of Copyrighted Works to Create New Technology?	TECHNOLOGY USAGE: Use of Copyrighted Works in Public Use of Technology?	FACTOR 1: Use of Copyrighted Works Had Further Purpose or Different Character?
American Broad. Co. v. Aereo, Inc., 573 U.S. 431 (2014).	No.	Yes, service enabled users to record TV shows using remote personal antennas and recording offered by online service.	[No court decision on fair use.]
Infinity Broad. v. Kirkwood, 150 F.3d 104 (2d Cir. 1998).	No.	Yes, retransmission of radio broadcasts over telephone.	No, service just "sell[s] access to unaltered radio broadcasts."
A&M Records, Inc. v. Napster, Inc., 239 F.3d 1004 (9th Cir. 2001).	No.	Yes, file-sharing copies online.	No, file sharing of music files "does not transform the copyrighted work."
Video Pipeline v. Buena Vista Home Ent., 342 F.3d 191 (3d Cir. 2003).	No.	Yes, service made "clip previews" of Disney movies sold to retail websites selling home videos.	No, clip previews substituted for Disney's movie trailers.
Capitol Records, LLC v. ReDigi, Inc., 910 F.3d 649 (2d Cir. 2018).	No.	Yes, service makes copies of music files to facilitate resales of them online, while attempting to ensure deletion of seller's copy.	No, service enables "resale of digital music files, which resales compete with sales of the same recorded music by the rights holder."
U.S. v. ASCAP, 599 F. Supp. 2d 415 (S.D.N.Y. 2009).	No.	Yes, wireless service provider planned to offer previews of ringtones of copyrighted music without license.	No, wireless carrier's preview of ringtones served same purpose.
Fox News Network, LLC v. TVEyes, Inc., 883 F.3d 169 (2d Cir. 2018)	Yes, to create a searchable database of TV and radio broadcasts	Yes, user views up to 10 minutes of recordings relevant to search topic.	Yes, but only modestly in allowing clients to time shift and to view what "they want at a time and place that is convenient.
Hachette Book Group v. Internet Archive, 115 F.4th 163 (2d Cir. 2024).	Yes, to create a searchable database of online library of books and other works, some of which are copyrighted	Yes, user given access to entirety of work.	No, service made "digital copies of the Works and distributes those copies to its users in full, for free."
UMG Recordings v. MP3.com, 92 F. Supp.2d 349 (S.D.N.Y. 2000).	No.	Yes, service copied "thousands of popular CDs in which plaintiffs held the copyrights, and, without authorization, copied their recordings onto its computer servers so as to be able to replay	No, service "simply repackages those recordings to facilitate their transmission through another medium."

		the recordings for its subscribers."	
Associated Press v. Meltwater U.S. Holdings, 931 F. Supp. 2d 537 (S.D.N.Y. 2013).	Yes, to create a searchable database for a news clipping service to provide clients with excerpts of news article	Yes, user receives 300-word excerpts of news articles relevant to searches, including email feeds	No, service "copies AP content in order to make money directly from the undiluted use of the copyrighted material" as a substitute of original works.
American Broad. Co. v. Aereo, Inc., 573 U.S. 431 (2014).	No.	Yes, service enabled users to record TV shows using remote personal antennas and recording offered by online service.	[No court decision on fair use.]

APPENDIX E: Map of U.S. Copyright Lawsuits v. AI Companies
Source: ChatGPT Is Eating the World