Mitchell L. Stoltz (SBN 291302)
mitch@eff.org
ELECTRONIC FRONTIER FOUNDATION
815 Eddy Street
San Francisco, CA 94109
Telephone: (415) 436-9333
Facsimile: (415) 436-9993

Attorney for Amicus Curiae
Electronic Frontier Foundation

UNITED STATES DISTRICT COURT

FOR THE NORTHERN DISTRICT OF CALIFORNIA (SAN FRANCISCO)

RICHARD KADREY, et al.,

Individual and Representative Plaintiffs,

v.

META PLATFORMS, INC. a Delaware Corporation,

Defendant.

Case No.: 23-cv-03417-VC

AMICUS CURIAE BRIEF OF THE ELECTRONIC FRONTIER FOUNDATION IN OPPOSITION TO PLAINTIFFS' MOTION FOR SUMMARY JUDGMENT

TABLE OF CONTENTS

TABLE OF AUTHORITIES ii

INTRODUCTION 1

ARGUMENT 1

A. The Application of Fair Use to LLM and Other AI Training Is an Important Issue of First Impression. 1

B. Fair Use Is a Fact-Sensitive Inquiry That Can't Be Shortcut. 5

C. The Fair Use Analysis Does Not Depend on How Meta Obtained the Books. 6

CONCLUSION 8

TABLE OF AUTHORITIES

Cases

A.V. ex rel. Vanderhye v. iParadigms, LLC,
562 F.3d 630 (4th Cir. 2009) 4

A&M Recs., Inc. v. Napster, Inc.,
239 F.3d 1004 (9th Cir. 2001) 6

Am. Inst. of Physics v. Winstead PC,
No. 3:12-CV-1230-M, 2013 WL 6242843 (N.D. Tex. Dec. 3, 2013) 7

Am. Soc'y for Testing & Materials v. Public.Resource.Org,
82 F.4th 1262 (D.C. Cir. 2023) 5

Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith,
598 U.S. 508 (2023) 5, 7

Atari Games Corp. v. Nintendo of Am., Inc.,
975 F.2d 832 (Fed. Cir. 1992) 6

Authors Guild v. Google, Inc.,
804 F.3d 202, 209 (2d Cir. 2015) 4, 5

Authors Guild, Inc. v. HathiTrust,
755 F.3d 87 (2d Cir. 2014) 4

BMG Music v. Gonzalez,
430 F.3d 888 (7th Cir. 2005) 6

Campbell v. Acuff-Rose Music, Inc.,
510 U.S. 569 (1994) 5, 7

Google LLC v. Oracle Am., Inc.,
593 U.S. 1 (2021) ..... *passim*

Harper & Row Publishers, Inc. v. Nation Enterprises,
471 U.S. 539 (1985) 7

In re DMCA § 512(h) Subpoena to Twitter, Inc.,
608 F. Supp. 3d 868 (N.D. Cal. 2022) 6

Kadrey v. Meta Platforms, Inc.,
No. 23-CV-03417-VC, 2023 WL 8039640 (N.D. Cal. Nov. 20, 2023) 5

Kelly v. Arriba Soft Corp.,
336 F.3d 811 (9th Cir. 2003) 4

NXIVM Corp. v. Ross Inst.,
364 F.3d 471 (2d Cir. 2004) 7

Online Pol'y Grp. v. Diebold, Inc.,
337 F. Supp. 2d 1195 (N.D. Cal. 2004) 5

Perfect 10, Inc. v. Amazon.com, Inc.,
508 F.3d 1146 (9th Cir. 2007) 4, 7

Sega Enterprises Ltd. v. Accolade, Inc.,
977 F.2d 1510 (9th Cir. 1992) 3, 4

Sony BMG Music Ent. v. Tenenbaum,
672 F. Supp. 2d 217 (D. Mass. 2009) 6

Sony Corp. of Am. v. Universal City Studios, Inc.,
464 U.S. 417 (1984) 2, 3

Thaler v. Perlmutter,
687 F. Supp. 3d 140 (D.D.C. 2023) 2

Triller Fight Club II LLC v. H3 Podcast,
No. LA CV21-03942 JAK (KSX), 2023 WL 118776048 (C.D. Cal. Sept. 15, 2023) 7

Twentieth Century Music Corp. v. Aiken,
422 U.S. 151 (1975) 3

United States v. Slater,
348 F.3d 666 (7th Cir. 2003) 6

Other Authorities

"How to train your large language model," *The Economist* (Mar. 13, 2024) 1

Anna Choi and Katelyn Xiaoying Mei, "What Are AI Hallucinations? Why AIs Sometimes Make Things Up," *The Conversation* (Mar. 21, 2025) 2

Jannis Douloumis, "The BitTorrent Revolution Never Ended," *Medium* (Dec. 4, 2024) 8

Lee Rainie, "Close Encounters of the AI Kind: The Increasingly Human-Like Way People Are Engaging with Language Models, *Elon University* (Mar. 2025) 2

Pamela Statz, "NASA Propels Downloads with BitTorrent," *Wired* (Jan. 27, 2006) 8

Pierre N. Leval, *Toward a Fair Use Standard*, 103 Harv. L. Rev. 1105 (1990) 7

INTRODUCTION

This case raises an important issue of first impression: whether internal copying necessary to train a generative artificial intelligence system is a non-infringing fair use. Judicial resolution of this issue could determine the future of this powerful new technology, including its capabilities, its costs, and whether its evolution will be shaped by the democratizing forces of the open market or the whims of an oligopoly.

This question deserves careful consideration on a full record. Plaintiffs' motion for partial summary judgment seeks to shortcut that careful consideration based on a legally tangential issue—how the defendant acquired the works—and a set of categorical presumptions that the Supreme Court has thoroughly rejected.

This Court should not allow the tail of Meta's alleged BitTorrent use to wag the dog of the important legal questions this case presents. Nor should it accept Plaintiffs' invitation to let hyperbole about BitTorrent and "unmitigated piracy" derail the thoughtful and fact-specific fair use analysis the law requires.

ARGUMENT

A. The Application of Fair Use to LLM and Other AI Training Is an Important Issue of First Impression.

How courts decide to apply fair use to training generative AI—currently an unsettled issue—will profoundly shape the future of this transformative technology. Developing a Large Language Model (LLM) like Meta's Llama requires "training" it on a very large volume of written material, allowing it to draw inferences and recognize patterns from the training data.[1] Once

[1] "How to train your large language model," *The Economist* (Mar. 13, 2024), at https://www.economist.com/science-and-technology/2024/03/13/how-to-train-your-large-language-model.

trained, the model provides original responses to user queries by calculating, based on the training materials, which words are most likely to appear in a sequence of text, and composing text based on those calculations. LLMs can generate text, images, music, video, or other sophisticated outputs based on a user's prompt. More than half of adults in the U.S. now use LLMs for a diverse array of tasks, including to search for information, brainstorm ideas, and to aid in composing everything from computer code to songs and poetry.[2]

Massive, high-quality data sets are essential to the training process, but not because the developers have any interest in fostering infringement. Rather, incomplete training data leads to poorer-functioning models more prone to error and bias.[3] Thus, training the AI on copyrighted works is an intermediate step necessary to build a general-purpose technology that provides information and produces original text at the user's request.

This general purpose technology is "approaching new frontiers in copyright" that "will prompt challenging questions" regarding the "scope of protection" that would best "incentivize creative works involving AI" and otherwise "promote science and the useful arts." *Thaler v. Perlmutter*, 687 F. Supp. 3d 140, 147, 149 (D.D.C. 2023), *aff'd*, No. 23-5233, 2025 WL 839178 (D.C. Cir. Mar. 18, 2025). Courts are "reluctan[t] to expand the protections afforded by the copyright…when major technological innovations alter the market for copyrighted materials." *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417, 431 (1984). In the face of "technological change," courts "must be circumspect in construing the scope of rights created by"

[2] Lee Rainie, "Close Encounters of the AI Kind: The Increasingly Human-Like Way People Are Engaging with Language Models, *Elon University* (Mar. 2025), https://imaginingthedigitalfuture.org/wp-content/uploads/2025/03/ITDF-LLM-User-Report-3-12-25.pdf.

[3] Anna Choi and Katelyn Xiaoying Mei, "What Are AI Hallucinations? Why AIs Sometimes Make Things Up," *The Conversation* (Mar. 21, 2025), https://theconversation.com/what-are-ai-hallucinations-why-ais-sometimes-make-things-up-242896.

the Copyright Act and must interpret the Act in light of copyright's purpose: stimulating "creativity for the general public good." *Id.* at 431-32 (quoting *Twentieth Century Music Corp. v. Aiken*, 422 U.S. 151, 156 (1975)) (internal quotation omitted).[4]

Fair use is one of the key provisions through which the law protects the public interest in promoting knowledge and learning in the face of technological change. Fair use is a "context-based check that can help keep a copyright monopoly within its lawful bounds," especially when "significant changes in technology" would otherwise expand rightsholders' power, to the public's detriment. *Google LLC v. Oracle Am., Inc.*, 593 U.S. 1, 22 (2021). The application of fair use to new technologies requires careful "judicial balancing, depending on relevant circumstances," including the "expressive and functional features" of the technology; the goal of promoting the creation and dissemination of new works; "the extent to which further protection creates unrelated or illegitimate harms in other markets or to the development of other products;" and the need to "prevent holders from using copyright to stifle innovation." *Id.* at 21-22. Courts "must avoid the temptation" to reflexively extend protections developed in "more traditional contexts" when considering new technologies that remain "relatively unexplored…in the world of copyright law"—or in other words, when trying to force "the proverbial square peg in[to] a round hole." *Sega Enterprises Ltd. v. Accolade, Inc*., 977 F.2d 1510, 1527 (9th Cir. 1992) (quotation omitted).

Closely analogous precedent suggests that using copyrighted works to train a generative AI model will be found to be fair use. Fair use can apply to the "wholesale copying" of creative

[4] *See also Sega Enterprises Ltd. v. Accolade, Inc*., 977 F.2d 1510, 1527 (9th Cir. 1992), *as amended* (Jan. 6, 1993) (giving "the copyright owner a *de facto* monopoly over…ideas and functional concepts" contained in computer code "defeats the fundamental purpose of the Copyright Act—to encourage the production of original works by protecting the expressive elements of those works while leaving the ideas, facts, and functional concepts in the public domain for others to build on.").

works as an "intermediate" or "necessary" step to develop similar technologies. *Google*, 593 U.S. at 22 (collecting cases). The Ninth Circuit has held that Google's reproduction of copyrighted images as "thumbnails" in search engine results was "highly transformative"—and thus fair use—because the search engine incorporated images "originally created to serve an entertainment, aesthetic, or informative function" into "an electronic reference tool" with immense public benefits. *Perfect 10, Inc. v. Amazon.com, Inc.*, 508 F.3d 1146, 1165-66 (9th Cir. 2007); *see also Kelly v. Arriba Soft Corp.*, 336 F.3d 811, 818-19 (9th Cir. 2003) (defendant's use of images as thumbnails in search engine results was "transformative" because it "serve[d] a different function than [plaintiff's] use—improving access to information on the internet versus artistic expression."). Fair use also applied to the copying of some 20 million books to create Google Books, a searchable database of books that locates and displays short excerpts of text from those books that contain the user's search terms. *Authors Guild v. Google, Inc.*, 804 F.3d 202, 209, 216-18 (2d Cir. 2015). The court held that this use had a "highly transformative purpose of identifying books of interest to the searcher" and was thus fair use, even though Google copied and indefinitely stored the text of entire books. *Id.* at 216-18; *see also Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87, 97 (2d Cir. 2014) ("[T]he creation of a full-text searchable database [of books] is a quintessentially transformative use."). Courts have also held that fair use applies to other uses of copyrighted materials to build products that aid in the creation of new works or enable new insights into existing works. *See, e.g., Sega*, 977 F.2d at 1527 (computer code used to create new video games); *A.V. ex rel. Vanderhye v. iParadigms, LLC*, 562 F.3d 630, 640 (4th Cir. 2009) (wholesale copying and storage of academic papers to build plagiarism detection software).

These precedents counsel skepticism of Plaintiffs' summary judgment motion. This Court has been "circumspect in construing the scope of rights created by a statute that never contemplated

such a calculus of interests," *Sony*, 464 U.S. at 417, by dismissing Plaintiffs' claims that training Meta's LLM on Plaintiffs' books makes the model itself and all of its outputs infringing derivatives of every work used in training. *Kadrey v. Meta Platforms, Inc*., No. 23-CV-03417-VC, 2023 WL 8039640, at *1 (N.D. Cal. Nov. 20, 2023). The same caution requires denying Plaintiffs' summary judgment motion, lest some rightsholder interests unduly "interfere with, not further, copyright's basic creativity objectives." *Google*, 593 U.S. at 39-40.

B. Fair Use Is a Fact-Sensitive Inquiry That Can't Be Shortcut.

A fundamental premise of Plaintiffs' motion is the idea that there are categories of acts that are always infringing as a matter of law. Not so. The fair use doctrine "permits courts to avoid rigid application of the copyright statute when, on occasion, it would stifle the very creativity which that law is designed to foster." *Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith*, 598 U.S. 508, 527 (2023) (citation omitted). Fair use requires courts to examine four statutory factors and weigh them together "in light of the purposes of copyright." *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 578 (1994). That analysis "is not to be simplified with bright-line rules." *Id.*

Accordingly, Plaintiffs' claim that "there exists a discrete set of infringing acts for which courts have held fair use cannot apply as a matter of law," Br. at 22, is incorrect. Commercial uses can be fair. *See Campbell*, 510 U.S. at 584. Valuable, uncompensated uses can be fair. *See Google*, 593 U.S. at 32. Copying thousands of books can be fair. *Authors Guild v. Google*, 804 F.3d at 225. Even making complete copies of entire works and making them available for unrestricted download over the internet can be fair use. *Am. Soc'y for Testing & Materials v. Public.Resource.Org*, 82 F.4th 1262, 1272 (D.C. Cir. 2023); *Online Pol'y Grp. v. Diebold, Inc.*, 337 F. Supp. 2d 1195, 1203 (N.D. Cal. 2004).

No court has held that peer-to-peer file sharing is categorically exempt from the four-factor fair use analysis. *Contra* Plaintiffs' Notice of Motion and Motion for Partial Summary Judgment (ECF No. 472), at 22-26. Nor could it, in light of the Supreme Court's repeated admonitions that fair use is a fact-sensitive inquiry. The cases Plaintiffs cite do not so hold. Some involved a perfunctory fair use defense. *United States v. Slater*, 348 F.3d 666, 669 (7th Cir. 2003); *Sony BMG Music Ent. v. Tenenbaum*, 672 F. Supp. 2d 217, 220 (D. Mass. 2009). Others raised only a personal use defense that was found to be insufficiently transformative. *BMG Music v. Gonzalez*, 430 F.3d 888, 891 (7th Cir. 2005); *A&M Recs., Inc. v. Napster, Inc.*, 239 F.3d 1004, 1015 (9th Cir. 2001). None of these cases dispensed with the four-factor analysis, as Plaintiffs suggest. This Court's opinion in *In re DMCA § 512(h) Subpoena to Twitter, Inc.*, 608 F. Supp. 3d 868, 879 (N.D. Cal. 2022) mentioned peer-to-peer filesharing only in dicta before proceeding to a thorough fair use analysis. Adopting an exemption from the four-factor analysis here, for "internet piracy," "P2P downloading," or something else, would contradict controlling precedent and undermine the purpose of the fair use doctrine as a safeguard for the application of copyright law to new technologies.

C. The Fair Use Analysis Does Not Depend on How Meta Obtained the Books.

Neither Meta's resort to sources like LibGen and Books3 nor its use of the BitTorrent protocol to obtain copies of books should impact its fair use defense.

First, there is no requirement that a would-be fair user must possess an authorized copy of a work. Fair use involves an overall balancing of the equities, including the potential public benefits of the use. *Google*, 593 U.S. at 35. In light of that directive, the "authorized copy" requirement cited by Plaintiffs that appears in *Atari Games Corp. v. Nintendo of Am., Inc.*, 975 F.2d 832, 843 (Fed. Cir. 1992) cannot be considered good law. The Ninth Circuit has implicitly

rejected this requirement. *See Triller Fight Club II LLC v. H3 Podcast*, No. LA CV21-03942 JAK (KSX), 2023 WL 11877604, at *8 (C.D. Cal. Sept. 15, 2023) (citing *Perfect 10*, 508 F.3d at 1164 n.8).

Second, although the Supreme Court said in *Harper & Row Publishers, Inc. v. Nation Enterprises* that "[f]air use presupposes good faith," 471 U.S. 539, 540 (1985), subsequent cases have made clear that the holding was narrow. *Harper & Row* involved a use that "had not merely the incidental effect but the *intended purpose* of supplanting the copyright holder's commercially valuable right of first publication." *NXIVM Corp. v. Ross Inst.,* 364 F.3d 471, 479 (2d Cir. 2004) (quoting *Harper & Row*, 471 U.S. at 562) (emphasis added). Other courts have read this holding to require, for example, obtaining a work through "breach of confidence or deception," *Am. Inst. of Physics v. Winstead PC*, No. 3:12-CV-1230-M, 2013 WL 6242843, at *12 (N.D. Tex. Dec. 3, 2013).

Third, in recent decisions the Supreme Court has been openly skeptical of the relevance of bad faith to the fair use inquiry. *Google*, 593 U.S. at 32 (citing *Campbell*, 510 U.S. at 585 n.18 (1994)). Its most recent fair use opinion ignores bad faith altogether in its analysis of the first statutory factor. *Warhol*, 598 U.S. at 508. The message is clear: Fair use, like copyright generally, "is not a privilege reserved for the well-behaved." Pierre N. Leval, *Toward a Fair Use Standard*, 103 Harv. L. Rev. 1105, 1126 (1990).

Finally, the use of BitTorrent to download books has no impact on the fair use analysis. BitTorrent is a protocol used for distributing large files efficiently across the internet. Like any information technology, BitTorrent has both legitimate and illegitimate uses. It has been used by

NASA to disseminate photos as part of the "Visible Earth" program.[5] It has also been used by the developers of widely used software including games and operating systems to distribute software updates.[6] The BitTorrent protocol and similar peer-to-peer technologies are also used for data backup, disaster response, and by the content delivery networks that underlie most legitimate distribution of digital entertainment.[7] Even if the Court considers Meta's good faith to be a relevant factor in the fair use analysis, using BitTorrent instead of other methods to obtain book texts does not show a lack of good faith. And the court should take care not to impose unnecessary legal risk on the use of important and widely used peer-to-peer technologies.

CONCLUSION

The issues raised by this case regarding the application of fair use to the internal and intermediate copying involved in training an AI system are too important to be determined by a tangential issue. The Court should deny Plaintiffs' summary judgment motion.[8]

Respectfully submitted,

DATED: April 3, 2025

/s/ Mitchell L. Stoltz

Mitchell L. Stoltz

ELECTRONIC FRONTIER FOUNDATION
815 Eddy Street
San Francisco, CA 94109
Telephone: (415) 436-9333
Facsimile: (415) 436-9993

Attorney for Amicus Curiae
Electronic Frontier Foundation

[5] Pamela Statz, "NASA Propels Downloads with BitTorrent," *Wired* (Jan. 27, 2006), at https://www.wired.com/2006/01/nasa-propels-downloads-with-bittorrent/.
[6] Jannis Douloumis, "The BitTorrent Revolution Never Ended," *Medium* (Dec. 4, 2024), at https://medium.com/@PowerUpSkills/bittorrent-a-breakdown-for-developers-and-tech-enthusiasts-it-is-all-in-here-860983bc4c88.
[7] *Id.*
[8] *Amicus* expresses no opinion regarding Meta's summary judgment motion.