UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

IN RE OPENAI, INC., COPYRIGHT INFRINGEMENT LITIGATION This document relates to: No. 1:23-cv-08292 (SHS) (OTW) No. 1:23-cv-10211 (SHS) (OTW) No. 1:25-cv-03482 (SHS) (OTW)	Case No. 25-md-3143-SHS-OTW Hon. Sidney H. Stein **PUBLIC REDACTED**

CLASS PLAINTIFFS' MEMORANDUM OF LAW IN OPPOSITION TO OPENAI DEFENDANTS' MOTION FOR SUMMARY JUDGMENT

TABLE OF CONTENTS

INTRODUCTION 1
FACTUAL BACKGROUND 3
LEGAL STANDARD 4
I. Summary Judgment 4
II. Principles of Fair Use 4
ARGUMENT 6
I. Use #1: OpenAI Has Failed to Meet Its Burden to Establish That Its Free Downloads of Plaintiffs' Books from LibGen and Books3 Was Fair Use 6
A. Factor One: OpenAI's Purpose in Pirating Millions of Books Was Substitutive, Not Transformative. 7
i. OpenAI's Free Downloading Is A Separate, Non-Transformative Use. 9
ii. OpenAI Cannot Establish LLM Training Was The Sole Use Of The Downloaded Books. 12
B. Factors Two and Three Favor Plaintiffs. 12
C. Factor Four: OpenAI's Conduct, If Widespread, Would Actual and Potential Markets for Plaintiffs' Works In Multiple Ways. 13
II. Use #2: OpenAI Has Failed to Meet Its Burden to Establish That Copying Plaintiffs' Works to Train Its LLMs Was Fair Use. 15
A. Factor One: OpenAI Copied Plaintiffs' Works For a Substitutive and Competitive Purpose Unconnected to Any "Transformative" Use. 18
i. The Competitive and Substitutive Nature of OpenAI's Use Precludes Summary Judgment in OpenAI's Favor. 18
ii. OpenAI's "Statistical Patterns" Account of LLM Training Is Disputed And Does Not Support Fair Use. 21
iii. OpenAI Had No Cognizable Justification for Copying Plaintiffs' Works. 25
iv. Kadrey and Bartz's Factor One Holdings Do Not Apply On This Record. 27
v. Google Books and the So-Called "Intermediate Copying" Cases Undermine, Rather Than Support, OpenAI's Fair Use Defense. 28
B. Factors Two and Three Favor Plaintiffs. 30
C. Factor Four Weighs Strongly Against Fair Use 31
i. OpenAI's Failure to Participate In The Existing and Growing AI-Training Data Licensing Market Is Dispositive On Factor Four. 34
ii. The Court Must Consider Harms From The Flood of AI-Generated Works Resulting From OpenAI's Use 36
D. Public Benefits 40
III. Use #3: OpenAI's Motion Entirely Fails to Address Its Distribution. 40
CONCLUSION 41

TABLE OF AUTHORITIES

Cases

A&M Records, Inc. v. Napster, Inc.,
239 F.3d 1004 (9th Cir. 2001) 6, 7, 9

A.V. ex rel. Vanderhye v. iParadigms, LLC,
562 F.3d 630 (4th Cir. 2009) 11

American Geophysical Union v. Texaco,
60 F.3d 913 (2d Cir. 1994)..... *passim*

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

Authors Guild, Inc. v. HathiTrust,
755 F.3d 87 (2d Cir. 2014).....10

Authors Guild v. Google, Inc.,
804 F.3d 202 (2d Cir. 2015)..... *passim*

Authors Guild v. HathiTrust,
1:11-cv-6531, Dkt. 113 10

Bartz v. Anthropic PBC,
787 F. Supp. 3d 1007 (N.D. Cal. 2025) ..... *passim*

Basic Books, Inc. v. Kinko's Graphics Corp.,
758 F. Supp. 1522 (S.D.N.Y. 1991).....29

Bill Graham Archives v. Dorling Kindersley Ltd.,
448 F.3d 605 (2d Cir. 2006).....14, 39

Blanch v. Koons,
467 F.3d 244 (2d Cir. 2006).....35

BMG Music v. Gonzalez,
430 F.3d 888 (7th Cir. 2005) 6, 15

Byrne v. Brit. Broad. Corp.,
132 F. Supp. 2d 229 (S.D.N.Y. 2001) (Stein, J.) 29

Campbell v. Acuff-Rose Music, Inc.,
510 U.S. 569 (1994).....17, 25, 35

Castle Rock Ent., Inc. v. Carol Publ'g Grp., Inc.,
150 F.3d 132 (2d Cir. 1998).....32

Fin. Guar. Ins. Co. v. Putnam Advisory Co., LLC,
No. 12 CIV. 7372 (AT), 2020 WL 264146 (S.D.N.Y. Jan. 17, 2020).....25, 40

Fox News Network, LLC v. TVEyes, Inc.,
883 F.3d 169 (2d Cir. 2018).....9, 16, 35

Google v. Oracle,
593 U.S. 1 (2021).....28, 30

Graham v. Prince,
2023 WL 3383029 (S.D.N.Y. May 11, 2023) (Stein, J.) ..... *passim*

Hachette Book Grp., Inc. v. Internet Archive,
115 F.4th 163 (2d Cir. 2024) ..... *passim*

Harper & Row Publishers, Inc. v. Nation Enters.,
471 U.S. 539 (1985).....13

Kadrey v. Meta,
788 F. Supp. 3d 1026 (N.D. Cal. 2025) ..... *passim*

Kalem Co. v. Harper Bros.,
222 U.S. 55 (1911).....6

Kelly v. Arriba,
336 F.3d 811 (9th Cir. 2003) 30

McGucken v. Newsweek LLC,
2022 WL 836786 (S.D.N.Y. Mar. 21, 2022) ..... *passim*

NXIVM Corp. v. Ross Institute,
364 F.3d 471 (2d Cir. 2004)..... *passim*

Oracle Am., Inc. v. Google LLC,
886 F.3d 1179 (Fed. Cir. 2018).....11, 30

Perfect 10 v. Amazon.com, Inc.
508 F.3d 1146 (9th Cir. 2007) 11, 30

Sega Enters. Ltd. v. Accolade, Inc.,
977 F.2d 1510 (9th Cir. 1992) 30, 38, 39

Sony Comput. Ent. Am., Inc. v. Bleem, LLC,
214 F.3d 1022 (9th Cir. 2000) 39

Sony Computer Entm't, Inc. v. Connectix Corp.,
203 F.3d 596 (9th Cir. 2000) 11, 21, 30

Thomson Reuters Enter. Ctr. GmbH v. ROSS Intel. Inc.,
2026 WL 2932557 (3d Cir. Sept. 29, 2026) ..... *passim*

Triangle Publ'ns, Inc. v. Knight-Ridder Newspapers, Inc.,
626 F.2d 1171 (5th Cir. 1980) 39

White v. West Pub. Corp.,
2014 WL 3385480 (S.D.N.Y. July 11, 2014) 35

Wright v. Warner Books, Inc.,
953 F.2d 731 (2d Cir. 1991) 4

Statutes

17 U.S.C. § 106 ..... 36, 40

17 U.S.C § 107 ..... 1, 36

17 U.S.C. § 107(4) ..... *passim*

Copyright Act 11, 37

Rules

Rule 56 4

Other Authorities

Black's Law Dictionary (12th Ed. 2024) 37

INTRODUCTION

OpenAI's Motion for Summary Judgment fails on the law and the facts. At its core, OpenAI attempts to excuse its rampant infringement by arguing that training a large language model is fair use. But copyright exists to provide "an economic incentive to create original works[.]" *Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith*, 598 U.S. 508, 535 (2023) ("*Warhol*"). As developed by the courts and codified by Congress, the fair use doctrine exists to avoid "rigid application of the copyright statute when, on occasion, it would stifle the very creativity which that law is designed to foster[,]" *id.* at 527 (internal quotations omitted), and for purposes "such as criticism, comment, news reporting, teaching[,] scholarship, or research[.]" 17 U.S.C § 107. In short, fair use is meant to amplify the purposes of copyright, not to destroy the very incentives copyright exists to protect.

OpenAI's arguments, if accepted, would decimate the book industry for authors and publishers alike and would allow unchecked piracy under the guise of fair use. OpenAI's arguments are wrong as matter of precedent, common sense, and fact—especially given the burden OpenAI bears on its fair use defense.

OpenAI's only real attempt to justify its behavior in downloading millions of books from LibGen and via Books3 is to maintain that its later allegedly fair use in training washes away any sin of pirated acquisition. This argument runs squarely counter to both *Warhol*, 598 U.S. at 528 and *American Geophysical Union v. Texaco*, 60 F.3d 913, 924 (2d Cir. 1994). These cases make clear that the court must evaluate any unlawful initial acquisition on its own terms and not in light of some subsequent act.

OpenAI also ignores the facts of piracy in *this case*. Here, OpenAI made a considered decision to pirate books instead of attempting to license them or otherwise obtain non-copyrighted works. They pirated *precisely because* piracy offered an easy way to avoid the costs, time, and

administrative burdens associated with legitimate acquisition. SUF¶138-181,283-304,519; RSUF¶104. But a company is not free to copy works without compensation simply because it would prefer to move faster and not pay. *Texaco*, 60 F.3d at 924. If a company can pirate without consequence by pointing to a later fair use, the same argument would allow a teenager to download pirated music to write a music review. Or a patron of the Pirate Bay to later trash the movie he downloaded on Metacritic. Put simply, writing a book review does not mean you can steal the book. On this basic point, even Sam Altman—along with a slew of OpenAI and Microsoft executives who now seek to run from their companies' conduct—agrees. SUF¶372,368-392.

The record further demonstrates that, contrary to its claims, OpenAI did not just acquire books for LLM training. Instead, it stored millions of books in internal repositories available to all employees. SUF¶244-261. It also distributed books to both Microsoft and third-party contractors—a use that OpenAI does not even try to justify in its motion. SUF¶571-596,607-611.

OpenAI's Motion separately fails because OpenAI's training was not fair use. Plaintiffs have moved for summary judgment because under the correct view of the law and the undisputed facts, OpenAI cannot prevail on its training fair use defense. *See* Pls. Mot. at 25-39. Since the Opening Motion, the Third Circuit's decision in *Thomson Reuters Enter. Ctr. GmbH v. ROSS Intel. Inc.*, 2026 WL 2932557, at *6, *8 (3d Cir. Sept. 29, 2026) ("*Ross*"), has only confirmed that exploiting copyrighted works without compensation to "create and optimize" an AI model capable of "commercial[ly] substitut[ing]" for the original works is not fair use. While *Ross* stated that it was not addressing generative AI models, its same rationale applies to OpenAI's conduct in gratuitously copying Plaintiffs' works to build a machine that would "substitute[e]" for these works in service of it (and Microsoft's) commercial gain. *See, e.g.,* SUF¶749. As Jack Clark put it right before the launch of GPT-3, "[a] ton of middle-class authors make their living via genre

fiction like sci-fi, romance, typically, DIY'd on Amazon via eBooks. ***This is exactly the sort of art product our systems are going to start being able to make soon***." Nelson Decl. Ex. 282 at -928 (emphasis added).

Even wrongly accepting OpenAI's legal premises, its Motion still fails for what are, at minimum, disputed issues of fact. Contemporaneous statements by OpenAI and Microsoft executives show that the model training process is not "transformative" "learning" but rather is a process of "compression" and "memorization" which encodes Plaintiffs' works precisely because they reduce training costs and improve the model's ability to generate substitutive outputs. *See infra* § II.A(a)-(b); ASUF¶45; RSUF¶23,49; SUF¶745-46,748, Pls. Mot. at 27. Moreover, evidence from Amazon, Plaintiffs' experts, and the Defendants themselves demonstrates both the active and growing market to license Plaintiffs' works for training along with the disastrous impacts from the flood of AI-generated books. *See infra* § II.C; SUF¶297,775-840,851,885-890; RSUF¶120,125-131; ASUF¶41-44; Pls. Mot. at 34-38.

OpenAI wants the unfettered ability to take the expressive content from books to create substitutive products for its own commercial purposes, all without any compensation to rights-holders. Such a position runs directly counter to the "economic incentive to create original works" protected by the Constitution. This Court should not contort the doctrine of fair use to allow a company to exploit Plaintiffs' expressive works without permission and for its own commercial purposes, especially where that same company recognized the threat these models posed to the livelihoods of authors and publishers.

FACTUAL BACKGROUND

Class Plaintiffs incorporate the facts contained in their Opening Motion, Dkt. 1848 ("Pls. Mot."), Statement of Undisputed Material Facts, Dkt. 1974 ("SUF"), and exhibits thereto, as well as the Responses to OpenAI's Statement of Facts ("RSUF") and Counterstatement of Fact

("ASUF") (and the similar filings addressing Microsoft's assertions) filed concurrently with this brief.

LEGAL STANDARD

I. Summary Judgment

As the party moving for summary judgment, OpenAI bears the burden to show that there "is no genuine dispute" as to any fact material to its fair use defense. *McGucken v. Newsweek LLC,* 2022 WL 836786, at *4 (S.D.N.Y. Mar. 21, 2022). Courts must "construe the facts in the light most favorable to the non-moving party and must resolve all ambiguities and draw all reasonable inferences against the movant." *Id.* (cleaned up). "When evaluating cross-motions for summary judgment, the Court reviews each party's motion on its own merits and draws all reasonable inferences against the party whose motion is under consideration." *Id.* (cleaned up). "The Second Circuit has [] urged that 'the fact-driven nature of the fair use determination suggests that a district court should be cautious in granting Rule 56 motions in this area[.]'" *Graham v. Prince*, 2023 WL 3383029, at *8 (S.D.N.Y. May 11, 2023) (Stein, J.) (quoting *Wright v. Warner Books, Inc.*, 953 F.2d 731, 735 (2d Cir. 1991)).

II. Principles of Fair Use

Application of the four (non-exclusive) fair use factors "always depends on consideration of the precise facts at hand." *Texaco*, 60 F.3d at 916 (citing *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 577 (1994)). And because "[t]he same copying may be fair when used for one purpose but not another," the fairness of *each* use must be considered separately. *Warhol*, 598 U.S. at 533. Even as to multiple "quoted passage[s]" contained in a "critical book review," courts must consider fair use as to "each" passage and cannot approve of any specific quotation merely because the "review, as a whole, serve[s] a different purpose than the book." *Id.* at 528, n.4.

Nearly two decades before *Warhol*, in *Texaco*, 60 F.3d 913, at 924, the Second Circuit made clear that a person who makes an unlawful copy of a work—there, a company's employees who made photocopies of journal articles in the course of performing research—"cannot gain fair use insulation" for its copying "simply because such copying is done by a company doing research." *Id.* What was true in *Texaco*, remains true after *Warhol*: no matter how "transformative" or valuable some later use may be, fair use as to the initial act of copying a book from a pirate site must stand, if at all, on its own. As *Bartz v. Anthropic PBC*, 787 F. Supp. 3d 1007, 1025 (N.D. Cal. 2025) explained:

> "There is no decision holding or requiring that pirating a book that could have been bought at a bookstore was reasonably necessary to writing a book review, conducting research on facts in the book, or creating an LLM. Such piracy of otherwise available copies is inherently, irredeemably infringing even if the pirated copies are immediately used for the transformative use and immediately discarded."

Warhol also shows that the central issue for fair use is "the problem of substitution—copyright's bête noire." 598 U.S. at 528. Such substitution encompasses and goes far beyond instances in which exact copies of a work are downloaded by or distributed to the public. As the Third Circuit recently recognized, in *Ross*, 2026 WL 2932557, at *6-10, the use of copyrighted materials to create a competing commercial product—even a product that itself was noninfringing—is a nontransformative, unfair use because it could compete in the original market for a rightsholder's work.

Against these principles, OpenAI suggests that the rights of those who create always give way to the interests of large technology companies. OpenAI's Memorandum of Law, Dkt. 1718 at 17-19 ("Mot."). Not so. Copyright exists to preserve the incentive for human authorship, and the fair use analysis is crafted to preserve that "economic incentive." *Warhol,* 598 U.S. at 531, 535. For this reason, courts have rejected fair use defenses over objections that liability would inhibit

technology progress, particularly where the claimed use exploits or profits from the protected expression in a copyrighted work and where this exploitation threatens to displace the market for human creation. From peer-to-peer networking software, *see A&M Records, Inc. v. Napster, Inc.*, 239 F.3d 1004, 1014-19 (9th Cir. 2001), to controlled digital lending libraries, *see Hachette Book Grp., Inc. v. Internet Archive*, 115 F.4th 163, 189 (2d Cir. 2024), to what were then-called "moving pictures," *see Kalem Co. v. Harper Bros.*, 222 U.S. 55, 61-63 (1911), courts have policed unauthorized copying in order to protect the economic incentive for human creation. *See Warhol*, 598 U.S. at 535. In each case—and despite claims of impossibility from advocates—the value of the market in question motivated erstwhile infringers and rightsholders to establish workable regimes of mutual benefit.

ARGUMENT

OpenAI made three distinct uses of Plaintiffs' works: (1) OpenAI downloaded the works—and millions of others—from pirate sites, stored the works in an internal repository, and made them available to all full-time employees and contractors; (2) OpenAI copied them in order to develop LLMs that could compete with Plaintiffs' Works; and (3) OpenAI distributed the works to Microsoft and other third parties. The court must assess each use separately, *Warhol*, 598 U.S. at 533. On none of these three uses has OpenAI carried its burden to establish fair use as a matter of law.

I. Use #1: OpenAI Has Failed to Meet Its Burden to Establish That Its Free Downloads of Plaintiffs' Books from LibGen and Books3 Was Fair Use.

For decades courts have rejected fair use when a party takes "for free something they would ordinarily have to buy." *Napster*, 239 F.3d at 1015 (quotation omitted); *BMG Music v. Gonzalez*, 430 F.3d 888, 890 (7th Cir. 2005). Multi-billion dollar technology companies should not be held to a different standard.

As Plaintiffs' Opening Motion explains, all four factors point inescapably to the conclusion that OpenAI's piracy was not fair use as a matter of law. Pls. Mot. at 19-25. Under factor one, the "purpose and character of" OpenAI's use was the same, substitutive, purpose as any other LibGen user: get books for free instead of paying for them. *See id.* at 19-22. As to factors two and three, Plaintiffs' books are at the heart of copyright and OpenAI copied them in their entirety. *Id.* at 22-23. And factor four, the market impact factor, weighs sharply against fair use for at least four reasons: OpenAI's free downloading deprived authors of book sales lost sales; OpenAI's downloading using the BitTorrent protocol facilitated the internet piracy of other shadow-library patrons via faster download speeds; OpenAI's conduct revitalized "dying" shadow libraries by firing an industry-wide starting gun which resulted in mass piracy; and would, if "unrestricted and widespread," decimate the legitimate market for Plaintiffs' books. *Id.* at 23-25.

OpenAI's motion mostly ignores these points, preferring instead to rest almost entirely on the argument that its later use of (some of) its millions of books to train LLMs immunizes it to liability for its piracy. *See* Mot. at 23-25. But that is not the law. Courts must assess fair use must with respect to each use. Even subsequent uses specifically identified for protection by the fair use provision—such as "research" or "news reporting"—do not excuse any initial, unlawful, and substitutive acquisition. *See Texaco*, 60 F.3d at 924.

A. Factor One: OpenAI's Purpose in Pirating Millions of Books Was Substitutive, Not Transformative.

As *Napster* made clear and *Bartz* confirmed, obtaining copyrighted works free from unauthorized sources that are otherwise available is non-transformative. *Napster*, 239 F.3d at 1015; *Bartz v. Anthropic PBC*, 787 F. Supp. 3d 1007, 1025 (N.D. Cal. 2025). A mountain of internal documents and testimony demonstrates that is precisely what happened here.

Starting in Fall 2018, OpenAI evaluated [redacted]

Just in the period from Fall 2018 to February 2019, OpenAI had lengthy discussions about [redacted]. Indeed, even after downloading millions of books from LibGen, OpenAI repeatedly [redacted]. SUF¶¶283-304.

But OpenAI rejected each of these opportunities and instead made a considered decision pirate them from LibGen, for free. Why?

Because doing so "offered an 'easy' way" to get the material it wanted, *Ross*, 2026 WL 2932557, at *7, because OpenAI thought licensing "discussions / negotiations" would be too "protracted," SUF¶288, because lawful acquisition might be a "multi-year, multi-person, multi-million [dollar] undertaking," SUF¶142, because taking the books from LibGen involved "no administrative burden," RSUF¶104, and because, at bottom, because OpenAI had a meeting with Bill Gates coming up and it wanted to attract billions of dollars of investment. SUF¶182-191,519.

In short, OpenAI "had many places from which it could have purchased books, but it preferred to steal them to avoid legal/practice/business slog." *Bartz*, 787 F. Supp. 3d at 1015 (cleaned up). The objective evidence thus demonstrates that the works OpenAI downloaded from LibGen and Books3 were not acquired by happenstance, but rather were the result of targeted,

intentional, and substitutive acts of piracy. As in *Bartz*, the "objective analysis here shows the initial copies were pirated . . . as a substitute for paid copies." 787 F. Supp. 3d at 1026.

OpenAI's conduct is no different than what teenagers did during the *Napster* era. In that context, courts held that downloading for free was not fair use, full stop. *See Napster*, 239 F.3d at 1015. As the Third Circuit recently confirmed, no copyright exception exists for artificial intelligence. *Ross*, 2026 WL 2932557. Certainly courts should hold a company worth over a trillion dollars to the same standard that it holds to private users of pirated content.

i. OpenAI's Free Downloading Is A Separate, Non-Transformative Use.

OpenAI argues that the allegedly transformative ultimate purpose of its downloading—LLM training—immunizes its decision to download copies of books free from LibGen and via Books3. Mot. at 23-25. That is wrong as a matter of Second Circuit law.

The Second Circuit's decision in *American Geophysical Union v. Texaco*, 60 F.3d 913, is controlling and dispositive of OpenAI's argument that its downstream use of Plaintiffs' works justifies its free acquisition of them. Like Defendants, Texaco argued that because its photocopying of journal articles facilitated research it was fair use. The Second Circuit disagreed, finding that Texaco's "extravagant" position would effectively exempt a wide variety of enterprises—from "newspaper[s]" to "school[s]"—from the copyright laws. *Id.* at 924; *see also Fox News Network, LLC v. TVEyes, Inc.*, 883 F.3d 169, 178 & n.4 (2d Cir. 2018). Indeed, OpenAI's conduct is even more egregious than Texaco's. Texaco, at least, paid for a subscription in the first place, and only infringed when their researchers made further copies to stock their personal office libraries.

Bartz reinforced that this same basic principle applies to LLM training: "You can't just bless yourself by saying I have a research purpose and, therefore, go and take any textbook you want. That would destroy the academic publishing market if that were the case." 787 F. Supp. 3d at 1025 (quoting counsel for Anthropic); *see Internet Archive,* 115 F.4th at 193 ("Were we to

approve IA's use of the Works, there would be little reason for consumers or libraries to pay Publishers for content they could access for free.").

Against this, OpenAI cites a series of cases, *see* Mot. at 24, but those cases largely undercut its claim. As *Bartz* explained, none of the cases cited by OpenAI ever "fathomed"—much less approved of—"gifting an 'artificial head start' to a fair user [] by treating *even the initial copy* as an intermediate one." 787 F. Supp. 3d at 1028 (emphasis added); *see id.* at 1026-29 (rejecting application of intermediate copying cases to unlawful downloads from LibGen).

In *Authors Guild v. Google, Inc.*, 804 F.3d 202 (2d Cir. 2015) ("*Google Books*") and its companion case *Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87 (2d Cir. 2014) ("*HathiTrust*"), Google and the consortium of "major research libraries" *paid for the* books that they used—exactly what OpenAI failed to do here. *See Authors Guild v. HathiTrust*, 1:11-cv-6531, Dkt. 113 ¶ 8 (S.D.N.Y. Jun. 29, 2012) (describing how libraries "spen[t] **tens [of] millions of dollars**" on book purchases); *Google Books*, 1:05-cv-08136, Dkt. 1036-1 at 107:9–108:3 (S.D.N.Y. July 27, 2012) (library spent roughly $20 million annually on book acquisition).

Nor does *Google Books* support OpenAI's argument that its downloading and subsequent use of certain books for training is a single "use." Mot. at 23-24. Even a glance at the opinion shows that the Second Circuit divided up the challenged uses into separate fair use analyses covering "Search and Snippet View," Google's "Exposure" of Plaintiffs' works to "Hacking," and "Google's Distribution of Digital Copies to Participant Libraries." *See Google Books* at 214-30; *see, also, e.g., Prince*, 2023 WL 3383029, at *7-23 (separately analyzing image's use in (1) art, (2) a gallery book, pamphlet, and billboard, (3) a Twitter compilation, and (4) an Instagram post).

OpenAI's reliance on *Kadrey v. Meta,* 788 F. Supp. 3d 1026 (N.D. Cal. 2025), is equally unavailing, as its reading of *Kadrey* squarely conflicts with controlling authority in this Circuit.

See Texaco, 60 F.3d at 924. *Texaco* and its progeny reject *Kadrey*'s suggestion that an overarching allegedly transformative purpose absolves all copies made of a work: a copyist cannot rely on a downstream "research" purpose to bless myriad unpaid and unauthorized copies along the way. *Texaco,* 60 F.3d at 924; *see Warhol*, 598 U.S. at 528-29 (requiring separate analysis of each instance of copying). In any event, the *Kadrey* Court later acknowledged that "Judge Alsup's opinion" treating Anthropic's free downloading as a separate wrong "makes a good point." *See Kadrey*, No. 23-cv-3417 (N.D. Cal Jul. 8, 2026), Dkt. 767.

OpenAI's other citations are offer no more support to its position. *See* Mot. at 24 n.4. In *Connectix*, and unlike OpenAI, the "engineers *purchased* a Sony PlayStation console" containing the copyrighted work. *Sony Computer Entm't, Inc. v. Connectix Corp.*, 203 F.3d 596, 601 (9th Cir. 2000); *see also, e.g.*, *A.V. ex rel. Vanderhye v. iParadigms, LLC*, 562 F.3d 630, 635 (4th Cir. 2009) (Plaintiffs "submitted their papers" to Defendant); *Oracle Am., Inc. v. Google LLC*, 886 F.3d 1179, 1187 (Fed. Cir. 2018) (the Java code Google copied had been "made available without charge under an open source license").

As for *Perfect 10 v. Amazon.com, Inc.* 508 F.3d 1146 (9th Cir. 2007), the copying was "in thumbnail form only," *Bartz,* 787 F. Supp. 3d at 1027, and the court held that "Google does not have a copy of the images for purposes of the Copyright Act," 508 F.3d at 1160. Nothing about *Perfect 10* mirrors OpenAI's conduct in downloading, organized, storing, and making internally available millions of books for whatever purposes its employees (and contractors) decided. *See Bartz,* 787 F. Supp. 3d at 1026-29; Pls. Mot. at 5-7, 11-12; SUF¶252-56,320; Shan Decl. Ex. A at ¶66; ASUF¶42.

OpenAI's initial downloading is separate from any later use. And that later use—no matter how arguably transformative—cannot absolve the initial acquisition. *Texaco*, 60 F.3d at 924.

ii. OpenAI Cannot Establish LLM Training Was The Sole Use Of The Downloaded Books.

The central factual premise of OpenAI's fair use defense to downloading is that OpenAI downloaded books from LibGen and Books3 solely to use those books for training. That fact is, at best, genuinely disputed and cannot be resolved in OpenAI's favor at summary judgment. Millions of books that OpenAI torrented from LibGen or downloaded from Books3 were not used in training, but instead put towards a separate use: to stock an internal central library available for any future purpose, with no access restrictions, and distributed to Microsoft and third-parties. Pls. Mot. at 5-7, 9-10, 11-12. This type of "archival" copying is akin to the bookshelf copies of academic articles that *Texaco* held was non-transformative. *See* 60 F.3d at 924.

OpenAI claims that *Bartz* is distinguishable because there is no evidence it sought to build a "general purpose library," Mot. at 25, but that is not the case. OpenAI maintained unauthorized copies of Plaintiffs books in central repositories, allowing its employees to reproduce them without restriction and maintain their own copies in personal storage spaces. This is precisely the type of central library infrastructure at issue in *Bartz.* There, Anthropic stored pirated books from LibGen. All the same features in *Bartz*—storage, searching tools, internal accessibility for whatever employees wished—are present here. *Compare* 787 F. Supp. 3d at 1016, 1026 *with* SUF¶252-56,313-14,320; ASUF¶50; Pls. Mot. at 9-11. And while OpenAI deleted a portion of these books in 2022, *see* Pls. Mot. at 11-12, nothing about the deletion alters the objective features of OpenAI's conduct for the four years prior to the deletion. Any lingering factual disputes over OpenAI's treatment of its pirated books only provide an additional reason to deny its motion for summary judgment.

B. Factors Two and Three Favor Plaintiffs.

Plaintiffs' Opening Motion explained why factors two and three favor Plaintiffs on OpenAI's piracy, *see* Pls. Mot. at 22-23, and Plaintiffs incorporate those arguments by reference.

C. Factor Four: OpenAI's Conduct, If Widespread, Would Actual and Potential Markets for Plaintiffs' Works In Multiple Ways.

Factor four considers "the effect of the use upon the potential market for or value of the copyrighted work." 17 U.S.C. § 107(4). The relevant impacts under factor four are not limited to the specific harms caused by the defendant's particular use but also include "adverse[] affect[s] [on] the *potential* market for the copyrighted work" were conduct like the defendants to become "widespread" *See Harper & Row Publishers, Inc. v. Nation Enters.*, 471 U.S. 539, 568 (1985) (quotation omitted).

As the party asserting the defense, OpenAI bears the burden of disproving market impact. *Internet Archive*, 115 F.4th at 193-95. Yet the portion of its motion addressing harms from its piracy contains no citation to any evidence, much less any evidence sufficient to carry its burden to disprove harm from its piracy or if that piracy were to become "unrestricted and widespread." *Id.* at 195; *see* Mot. at 35-36. This failure of proof by itself is dispositive.

By contrast, the evidence in this record precludes weighing the fourth factor in OpenAI's favor for three reasons. *First,* widespread conduct like OpenAI's here—*i.e.*, obtaining books for free in lieu of paying—would have an obvious impact on the market for Plaintiffs' works in the form of lost sales. Piracy annually costs hundreds of millions of dollars in lost sales. *See* Seeley Decl. Ex. A at ¶¶53-62.

Second, a reasonable juror could find that OpenAI's conduct—and widespread conduct like it—helped sustain shadow libraries like LibGen, and harmed authors by "normaliz[ing] piracy" throughout the AI industry and "encourag[ing] others to use shadow libraries." *Id.* at § VIII, ¶63. Shadow libraries themselves have reported that they "were dying," but then were

revitalized when "[v]irtually all major companies building LLMs contacted us to train on our data." *Id.* at ¶65, 63-76. *See also Bartz*, 787 F. Supp. 3d at 1015 (two of the OpenAI employees here later did the same thing at Anthropic); SUF¶396 (Leo Gao—now an OAI employee—created the Books3 set—pirated material he referred to as "yarrharr" data—because he wanted to keep up with OpenAI). This provides an independent basis to deny summary judgment. *See Kadrey*, 788 F. Supp. 3d at 1047 (where the plaintiff puts forth evidence that the defendant's piracy "benefitted those who created the [shadow] libraries and thus supported and perpetuated their unauthorized copying and distribution of copyrighted works . . . this sort of peer-to-peer file-sharing will constitute copyright infringement"); *see id.* at 1057-58.

Third, OpenAI's torrenting of LibGen facilitated the pirated downloads of other third-party users of the same torrent files—i.e., the same copyrighted works. Plaintiffs' expert Dr. David Choffnes reviewed OpenAI's torrenting source code and confirmed that the settings for OpenAI's torrenting would have directly benefited other LibGen users by increasing their download speeds, and thus enabled others to download Plaintiffs' works free. *See* Choffnes Decl. Ex. B at ¶33-43. *See Google Books*, 804 F.3d at 227 (concluding that evidence that a secondary use creates an "unreasonable expos[ure]" to piracy, even if unintentional, is a "substantial rebuttal to a secondary user's claim of fair use").

Lacking evidence, OpenAI claims that downloading books from LibGen instead of "buy[ing] the books . . . demonstrates nothing" because there is no "cognizable market harm beyond the mere act of copying." Mot. at 35-36. But this ignores the record evidence that OpenAI's conduct led to and facilitated other third parties' ability to access Plaintiffs works, a theory that *Google Books* recognized was "sound." *Google Books*, 804 F.3d at 227. Furthermore, OpenAI's citation to *Bill Graham Archives v. Dorling Kindersley Ltd.,* 448 F.3d 605, 614 (2d Cir. 2006),

concerns licensing fees (addressed below), not the initial acquisition itself. Instead, as the Second Circuit explained in *Internet Archive,* 115 F.4th at 195, "each digital book" OpenAI copied for free "deprives [] authors of the revenues due to them as compensation for their unique creations." *See also BMG Music*, 430 F.3d at 890-91 ("[A]ll 1,000+ of [defendant's] downloads . . . . undermined the means by which authors seek to profit.").

OpenAI also claims that there is "no evidence that [it] would have purchased new books from Plaintiffs." Mot. at 36. But the record demonstrates that OpenAI's at-the-time alternative to piracy was to purchase or otherwise license the books from [redacted] or publishers. *See supra* § I.A; Pls. Mot. 5-7; SUF¶176,182-191,288; RSUF¶10.

Finally, OpenAI claims that harms from a single lost sale are not "significant" enough and asserts, without evidence, that the "widespread conduct" inquiry would not change the result. Mot. at 36. But OpenAI bears the burden to disprove harm as to both its piracy *and* harms from "unrestricted and widespread" conduct of the same kind. It is "self-evident" that if pirating Plaintiffs' works without compensation "were to become widespread, it would adversely affect [the] markets for the Works in Suit." *Internet Archive,* 115 F.4th at 189, 193, 195.

II. Use #2: OpenAI Has Failed to Meet Its Burden to Establish That Copying Plaintiffs' Works to Train Its LLMs Was Fair Use.

As set forth in Plaintiffs' Opening Motion, OpenAI's copying of Plaintiffs' works to train LLMs was not fair use as a matter of law. *See* Pls. Mot. at 25-38. Internal documents, testimony from OpenAI and Microsoft staff, and expert testimony all confirm that OpenAI exploited Plaintiffs' expressive content for the non-transformative purpose of developing a commercial product that would compete with Plaintiffs. In particular, OpenAI trained on books to make the model better at writing books. *See id.* at 25-29. And while OpenAI claims that its models are capable of doing vastly different things unrelated to generating competing content, it admits

Plaintiffs' works had no impact on the models' ability to tell a user how to squeeze in a workout during your child's naptime. *See id.* at 29-32. In other words, Plaintiffs' books were not necessary to train the model to do what OpenAI claims is transformative. They *were*, however, valuable in improving the model's capability to generate prose that could sell on Amazon. *See id.* And as to the fourth factor, licenses signed by both OpenAI and Microsoft and uncontested [redacted] data all demonstrate the significant harm to traditional and licensing markets from OpenAI's unauthorized use. *See id.* at 33-38

Against this, OpenAI seeks to collapse all of its uses and the entirety of the fair use inquiry down to one oversimplified question: is training an LLM transformative? If so, according to OpenAI, then the facts and the remaining fair use factors do not matter, and the court cannot even consider the substantial harms from OpenAI's conduct. *See* Mot. at 20-23, 29-38. That is not the law. *See, e.g., Texaco,* 60 F.3d at 924-25; *TVEyes*, 883 F.3d at 179-180.

OpenAI rests this reduction on two points: (1) that LLM training is transformative because LLMs "learn" only "general concepts, and facts about the world" from Plaintiffs' books and the resulting LLMs are "general purpose" tools, Mot. at 20-21; and (2) any negative impacts on the market for Plaintiffs' books or the licensing market are illusory and not cognizable because OpenAI's use was transformative and OpenAI does not (always) reproduce the text of works to end users when prompted, *id.* at 29-38. Both claims are factually and legally incorrect.

Legally, OpenAI cannot prevail on summary judgment because, as to factor one, its use was non-transformative, competitive, and substitutive, and it lacks any justification connecting its exploitation of Plaintiffs' works to any transformative purpose. *Ross*, 2026 WL 2932557 at *6-7. As to factor four, the harms to already-existing licensing markets are cognizable even if OpenAI used Plaintiffs' works for "research" or "learning," *Texaco*, 60 F.3d at 924-25, *Ross*, 2026 WL

2932557 at *6-7, *Campbell*, 510 U.S. at 590-94, and the harms to Plaintiffs' traditional markets from AI-generated dilution are "effect[s]" of OpenAI's training use on the "market for or value of the copyrighted work" which must be considered and disproven for OpenAI to prevail at this stage. *See* 17 U.S.C. § 107(4); *Ross*, 2026 WL 2932557 at *8-9.

Factually, the Court cannot grant OpenAI summary judgment on training for three reasons.

First, OpenAI's entire account of the training process—and on which it rests its entire transformativeness argument—is contested by the record and Plaintiffs' experts. The record shows that OpenAI targeted Plaintiffs' books for substitution and that OpenAI trained its LLMs on Plaintiffs' books instead of public domain books precisely because doing so reduced the costs of training and made its LLMs better at producing outputs that were more similar to Plaintiffs' books. *See infra* § II.A(a) The record—including statements from OpenAI's executives—demonstrates that training is not a process of learning "unprotectable linguistic rules," Mot. at 20, but rather a process of compression which encodes Plaintiffs' expressive content into the model. *See infra* § II.A(b).

Second, genuine factual disputes over the existence and development of the licensing market for AI training also preclude summary judgment. Licenses signed by OpenAI and Microsoft to use books in AI training, testimony and documents from OpenAI and Microsoft's executives and employees, and Plaintiffs' expert analysis demonstrates that this market exists and is growing. *See infra* § II.C.

Third, OpenAI cannot carry its burden to show there is no genuine dispute of material fact as to the harms in diluting the market for Plaintiffs' books. Record evidence from [redacted], Plaintiffs' experts, and Plaintiffs themselves documents that OpenAI's models are capable of producing substitutive outputs in minutes and that the flood of AI-generated books enabled by

OpenAI's training use is negatively impacting the market for human authored books. *See supra* § II.C. OpenAI's evidence against this obvious and growing reality consists of expert analysis which did not even evaluate impacts to the e-book market or account for the [redacted], RSUF¶125-26, and does not suffice to carry its burden on factor four.

A. Factor One: OpenAI Copied Plaintiffs' Works For a Substitutive and Competitive Purpose Unconnected to Any "Transformative" Use.

OpenAI bears the burden on factor one of the fair use analysis—the purpose and character of the use. *See Texaco*, 60 F.3d at 918. But OpenAI cannot carry this burden because its purpose was competitive and substitutive; because its account of the technical nature of the training process—its implementation and aims—is contradicted by the record; and because OpenAI cannot show *any* justification connecting its copying of Plaintiffs' works to the purportedly transformative character of its use.

i. The Competitive and Substitutive Nature of OpenAI's Use Precludes Summary Judgment in OpenAI's Favor.

Under *Warhol*, substitution is a central question for the first factor. 598 U.S. at 528. The Third Circuit's recent opinion in *Thomson Reuters v. Ross Intelligence* is instructive on the contours of the first factor analysis. 2026 WL 2932557 at *6-7, In *Ross*, the Court analyzed copying of Westlaw headnotes by a competing legal research startup, Ross Intelligence. Ross had hired a vendor to prepare legal memos using (and copying) Westlaw headnotes, converted the memos into a "machine-readable format," and used them to "train the ROSS AI program to identify patterns regarding which judicial opinion passages respond well to legal questions." *Id.* at *2. Ross "ran ads directly comparing its platform to Westlaw" and its executives "hoped ROSS would serve as a substitute for Westlaw, and some law firms switched from using Westlaw." *Id.* at *3. The resulting tool displayed none of the underlying copyrighted headnotes to end users. *Id.* at *2.

The court found factor one weighed against fair use principally on two grounds. First, because both the plaintiff and defendant used the material at issue to sell similar end products to consumers—legal research tools—both uses shared the "same ultimate purpose" making the defendant's use "minimally transformative, at best"—even though Ross's competing commercial product did not distribute copies of the underlying copyrighted headnotes. *Id.* at *6. Second, the *Ross* court rejected the defendant's intermediate-copying-based justification arguments by pointing out that those cases did not apply where a defendant had no need to copy the protected material but only did so out of "ease," which, "[u]nlike necessity . . . is not a justification for copying." *Id.* at *7.

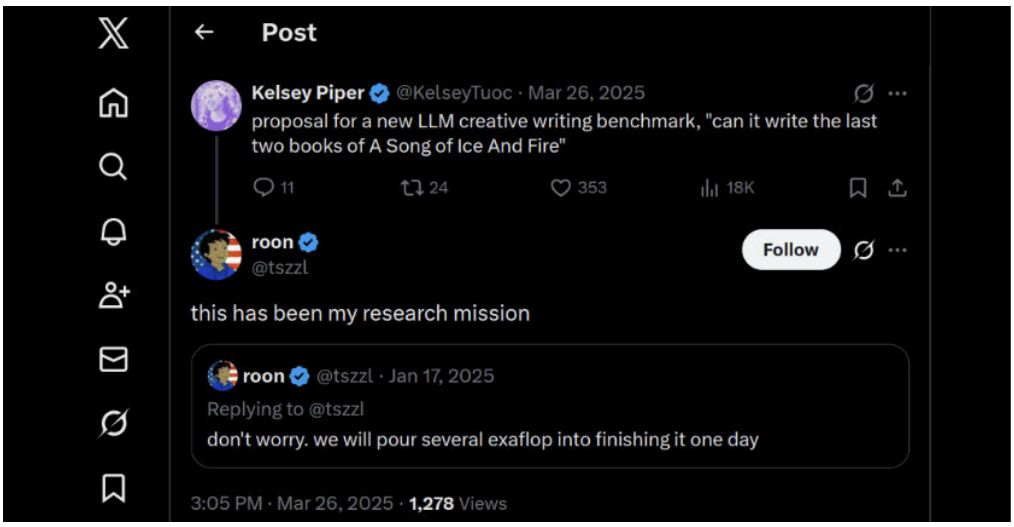
Here, the record is rife with evidence that OpenAI's copying shared the same ultimate purpose as Plaintiffs' works and that OpenAI used Plaintiffs works for ease and not out of necessity. *See* Pls. Mot. at 25-32.

First, OpenAI's executives repeatedly admitted internally that their exploitation of human authorship was designed to result in a commercial product to replace it entirely. One OpenAI executive admitted that their products would "ultimately substitute[] for the practiced, deeply personal labor of the creative communities of the world" and in particular "authors . . . on Amazon." SUF¶749-50. A novelist employed by OpenAI remarked that, after seeing early model outputs, his "days [were] numbered" as soon the models would "write the books!" SUF¶711; SUF¶730; Nelson Decl. Ex. 282 at -928 ("A ton of middle-class authors make their living via genre fiction like sci-fi, romance, typically, DIY'd on Amazon via eBooks. This is exactly the sort of art product our systems are going to start being able to make soon.").

That OpenAI's use would substitute for human authors was equally obvious outside the company. LinkedIn co-founder and Microsoft Board Member Reid Hoffman co-authored a book

with GPT-4 just a month following the model's public release and sent it separately to Microsoft's CEO Satya Nadella and Sam Altman. ASUF¶14-17. Leo Gao testified that it "didn't take long" after the launch of "GPT-3" to understand that "writing" was going to be a key use case. ASUF¶11-12. Tarun Gogineni similarly wrote that, beginning with GPT-3, OpenAI's models were already "very very good at writing stories." SUF¶717.

OpenAI's substitutive purpose was both general, in that it targeted book writing across the board, and specific, in that it targeted Plaintiffs in this case. Two years after George R. R. Martin sued OpenAI in this lawsuit, Gogineni—the head OpenAI's model writing quality team—posted publicly that his "research mission" was to have OpenAI's LLMs to a "write the last two books of [George R. R. Martin's] *A Song of Ice And Fire*.'"



SUF¶727.

OpenAI and Microsoft targeted Plaintiffs' market in particular: "writing" was "one of the top use cases on ChatGPT," accounting "for [redacted]" SUF¶723. Both Microsoft and OpenAI have extensively marketed LLMs for writing—including in pages like "Write your book with AI Help." SUF¶740-44. Against this bevy of evidence and at

summary judgment, OpenAI cannot establish that its use was transformative, rather than substitutive.

OpenAI argues that the problem of substitution is implicated only where a defendant distributes copies of a plaintiff's works. *See* Mot. at 20-23, 27-34. Not so. No case limits copyright infringement to violations of the distribution right. *Ross* reinforces this conclusion, as the court found a substitutive purpose even though the Westlaw headnotes were copied merely as an "intermediate step" to create a competing commercial product. 2026 WL 2932557 at *6-7. True, *Ross* is not a case about generative artificial intelligence, *id.* at *6 n.7, but neither is *Google Books*, *Connectix*, or *Texaco*. *Ross* refutes Defendants' central legal premise that exploiting protected expression as an intermediate step to create a competing commercial product is fair use.

Second, OpenAI falls short of its burden on summary judgment for an independent reason: It cannot establish a justification for its copying. *See Warhol*, 598 U.S. at 531, 547 & n.21. OpenAI had plenty of other options for training material, including public domain books or licensed data, and synthetic data. *See, e.g.,* SUF¶283-303,484-485. Copyrighted books like Plaintiffs' works were not necessary, but they made the training process cheaper and singularly improved the model's ability to generative substitutive content (i.e., higher-quality prose). SUF¶153,297,301-03,467-85,706-13,745-49,784. In fact, OpenAI employees conducted experiments to test the effect of replacing copyrighted LibGen books with public domain books. The results: models functioned properly when trained on the public domain, but it was easier and cheaper to use copyrighted books. RSUF¶23; SUF¶283-303. Because "ease is not a justification for copying," *Ross*, 2026 WL2932557 at *7, OpenAI cannot establish its justification on the first factor beyond a dispute of material fact—and that is independently dispositive of its summary judgment motion.

ii. OpenAI's "Statistical Patterns" Account of LLM Training Is Disputed And Does Not Support Fair Use.

OpenAI grounds its argument entirely on its account of the LLM training process as a process of learning "broad statistical patterns about human language" as a whole. *See* Mot. at 20. None of this matters. Even if OpenAI were correct as a factual matter, copying to generate patterns of language to create a competing commercial product is not fair use. *Supra* § II.A(a). But OpenAI's argument fails for another reason: At best, OpenAI's account of the training process is premised a serious of hotly contested material facts on which its entire transformativeness theory depends. To the extent that the Court must resolve the parties' dueling accounts of the training process—as OpenAI's motion puts at issue—this cannot be resolved on summary judgment. *See McGucken,* 2022 WL 836786, at *11 (denying summary judgment on fair use because "genuine disputes of material fact" meant that "a jury could reasonably find in favor of either party on … transformative[ness]").

First, evidence from Plaintiffs' experts directly contests OpenAI's claim that the training process optimizes the model to "learn" "broad statistical patterns about human language" or other "unprotected facts"—the factual assertions on which OpenAI's claim to transformativeness rests. Mot. at 20.

Instead, as Plaintiffs' expert Dr. Ben Zhao explains, the loss function—the same basic mathematical measure employed by compression software—is the technical core of the LLM training process and optimizes the model to faithfully reproduce the training data. Rather than a model of all "human language," a "large language model" is a model of *its training dataset*—and only its training dataset—created through a process which optimizes encoding the expressive content of Plaintiffs' works into the model. SUF¶720; RSUF¶¶23,37-49. As Dr. Zhao demonstrates, the so-called "tests" which OpenAI identifies as its "core metric" and on which it rests its claims of generalization, Mot. at 8, 21, do not reliably distinguish between a model which

memorizes and a model which generalizes. As Plaintiffs' expert Shawn Shan demonstrated in his output log and regurgitation testing, OpenAI's models produce verbatim passages from Plaintiffs' books—books which constituted a tiny fraction of the full training data—because those expressive elements were retained in the model's weights and not (as OpenAI posits) discarded in favor of abstract "patterns." *See, e.g.,* Mot. at 26; RSUF¶49,56,84; Shan Decl. Ex. A at ¶19-21, 140-150; Shan Decl. Ex. B at ¶14-51; Zhao Decl. Ex. B ¶72-76.[1]

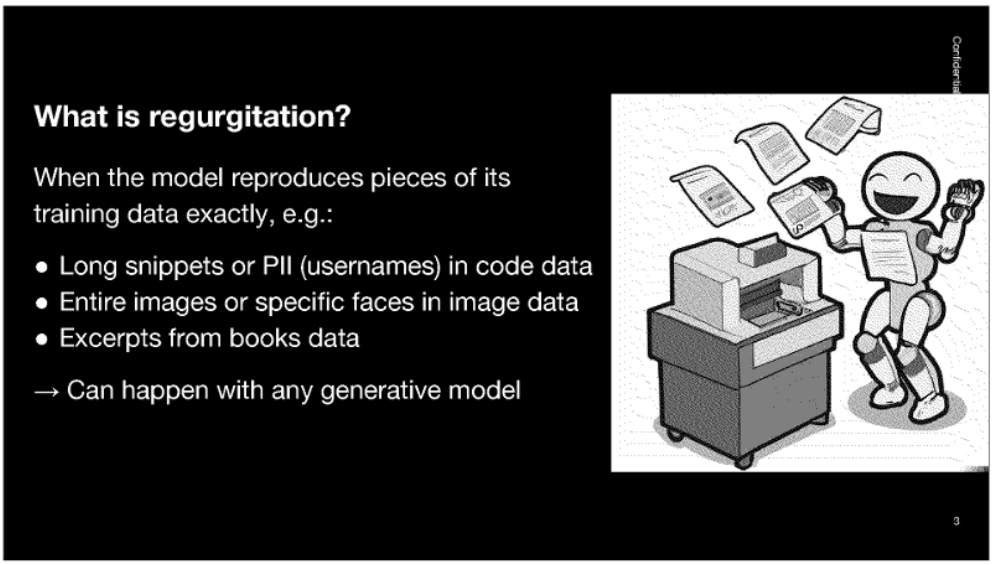
Second, statements by OpenAI's most senior executives confirm that OpenAI's characterization of the training process as a "transformative" process of generalized "learning," Mot. at 11—rather than an attempt to compress the data on which it trained—is at best a hotly contested question of fact. *See* RSUF¶49. As OpenAI co-founder Ilya Sutskever wrote in a memo sent to Bill Gates as part of OpenAI's efforts to induce Microsoft investments: the "***training of GPT-2-type models is compression*** *...* ***A GPT-type model is maximizing the likelihood of all its training data, which happens to be almost precisely equivalent to compressing it.*** … [it] ***compresses only the training data****...*" ASUF¶¶57-58.

Third, contemporaneous statements by OpenAI employees similarly contradict OpenAI's claim that the "goal" of the training process is to "distill broad statistical patterns" in service of "general purpose software." Mot. at 20. Instead, they openly affirmed that the "training objective" makes models "want to regurgitate" their training data. RSUF¶44. As one OpenAI employee put it:

> We train our networks to memorize the training data — that's their objective . . . our networks have lots of parameters and we train them to memorize the dataset . . *Id.*

[1] Plaintiffs are not claiming that every piece of training data is perfectly captured in reproducible form inside OpenAI's LLMs. Rather that training aims at and optimizes for memorization and compression, instead of amorphous "understanding" of human language. Mot. at 3.

In an internal presentation, OpenAI's Alec Radford acknowledged "[c]urrent techniques are closer to memorization than generalization." RSUF¶58. And Ben Mann testified that "[l]arge models appear to be capable of memorizing data after seeing it only once." RSUF¶64. In another presentation, OpenAI explained that anti-regurgitation efforts are "fighting against *the exact training objective* here."



We're fighting against the exact training objective here

ASUF¶4.

Fourth, the evidence in the record directly contradicts OpenAI's claims that it did not seek to "repackage Plaintiffs' works and display them to a human audience," given the efforts it claims to have made at anti-regurgitation. *See* Mot. at 20. Research from Microsoft showed that, even relative to other LLM providers, OpenAI's LLMs memorized and regurgitated far more. *See* RSUF¶73,75. Despite this, OpenAI released two of the at-issue models—GPT-3 and GPT-3.5—and ChatGPT itself without *any* anti-regurgitation safeguards, and acknowledged that it also released subsequent models without solving regurgitation. RSUF¶65; SUF¶564. Plaintiffs' Expert

Dr. Shawn Shan documented cases in which users turned to OpenAI's models to write summaries and unauthorized sequels to Plaintiffs' works which, in turn, reproduced portions of Plaintiffs' works. Shan Decl. Ex B ¶31. In light of the evidence above demonstrating that OpenAI trained on copyrighted books because it made their models better at producing outputs resembling those books, *see supra* at 16, factual disputes preclude finding that OpenAI did not seek to have its models "repackage" Plaintiffs' works to its users, Mot. at 20.

In sum, record evidence from Plaintiffs' experts and OpenAI's own executives and employees directly contests OpenAI's claims that the training process is "transformative" because models "distill broad statistical patterns about human language." Mot. at 20.[2] Any factual disputes regarding the nature of the LLM training process must be heard by a jury and decided at trial. *See McGucken,* 2022 WL 836786, at *11.

iii. OpenAI Had No Cognizable Justification for Copying Plaintiffs' Works.

As set forth in Plaintiffs' Opening Motion, *Campbell*, 510 U.S. at 580-81, and *Warhol*, 598 U.S. at 531, 547 & n.21, make clear that OpenAI must show a cognizable justification for its copying to prevail on factor one. That requires showing that its copying was "reasonably necessary" to its purportedly transformative purpose, *Warhol,* 598 U.S. at 532, particularly given the evidence of substitution and substitutive purpose here. *See supra* §II.A(a). OpenAI (and Microsoft) cannot do so where its own experts are unequivocal that copying Plaintiffs' works was entirely unnecessary to the achievement of OpenAI's purported purpose of training a "generalized" or "general use" LLM. *See* Pls. Mot. at 30-31; SUF¶¶496-516. Instead the record shows that

[2] In a footnote, OpenAI suggests that Plaintiffs do not have any output case, despite the Court expressly rejecting that argument. Dkt. 701 at 17-18. OpenAI's argument is waived. *See, e.g., Fin. Guar. Ins. Co. v. Putnam Advisory Co., LLC,* 2020 WL 264146, at *2 (S.D.N.Y. Jan. 17, 2020). In any case, OpenAI has not sought summary judgment on outputs containing Plaintiffs' content. The Court can deny OpenAI's motion on this ground.

OpenAI copied Plaintiffs' books because it was quick, easy, cheap, and enhanced their ability to sell substitutive outputs to their customers. *See supra* § II.A(a). None of that is transformative—and certainly not as a matter of law.

OpenAI's in response boils down to: we needed a lot of material, so we can copy anything and everything we want without compensation or permission. *See* Mot. at 27-28. That is *not* a justification under factor one because it does not connect the particular copying of Plaintiffs' works at issue here to any cognizable transformative purpose. *See* Pls. Mot. at 29-30. The question is not whether copying was "helpful"—as *Warhol*, 598 U.S. at 547, noted it "often is"—or whether customers enjoyed the secondary use—they often do, especially at the lower prices infringers charge—but instead whether the copying was reasonably necessary to achieve the transformative purpose. Where a defendant "could have used any number of other [works,] . . . it suggests that [the defendant's] commentary has no critical bearing on [the original]" and thus that factor one favors the copyright owner. *See Prince*, 2023 WL 3383029, at *13 (cleaned up). Here the record unequivocally establishes that OpenAI "could have used any number of other [works]," *id.*, including public domain books, *see* SUF¶301-03, and still achieved whatever transformative purpose and general use capabilities it now claims. Indeed, Sam Altman testified that OpenAI has "made extraordinary progress on being able to train models on very small amounts of data with very large amounts of compute." ASUF¶65.

As *Ross,* 2026 WL 2932557, at *7, explained when addressing the application of the first factor to similarly gratuitous copying by an AI company seeking—like OpenAI—to develop a competitor for cheap:

> Here, ROSS does not need to copy Thomson Reuters's headnotes to access the underlying unprotected information. …. It chose not to do so because copying the headnotes offered an "easy" way to create its training memos. … Unlike necessity, ease is not a justification for copying.

iv. *Kadrey* and *Bartz's* Factor One Holdings Do Not Apply On This Record.

All parties cite *Kadrey v. Meta,* 788 F. Supp. 3d 1026, and *Bartz v. Anthropic,* 787 F. Supp. 3d 1007, throughout their briefing, and Defendants cite both cases for the premise that LLM training is transformative. Both case's abbreviated records were different than this case.

In *Kadrey*, there was no "indication that … Meta developed [its LLM] with the purpose of enabling it to create books that compete with the plaintiffs'[.]" 788 F. Supp. 3d at 1045. As Judge Chhabria recognized in *Kadrey*, when an LLM is "designed to be used to create works . . . that competed with the originals without being substantially similar," the LLM is likely "less transformative." *Id.* n.6. The record here shows that, even if OpenAI sought to have its LLM perform a variety of tasks, OpenAI in fact did design the at-issue LLMs to compete with Plaintiffs' originals. This case presents exactly the "less transformative" situation contemplated by *Kadrey*, and, at most, confirms why a jury must resolve the disputed factual questions about the role of Plaintiffs' books in the training process.

As to *Bartz*, "Anthropic's LLMs trained upon works not to race ahead and replicate or supplant them," but to "create something different." 787 F. Supp. 3d at 1022. But, again, the record here shows that, whatever other functions it sought, OpenAI used Plaintiffs' works in LLM training without authorization or compensation specifically because it allowed them to move faster, save money, and create an LLM capable of replicating and supplanting Plaintiffs' works. *See supra* § I.A.

Bartz's contrary justification holding—that "[b]ecause using so many works was reasonably necessary, using any one work for actually training LLMs was about as reasonable as the next" 787 F. Supp. 3d at 1030—cited no precedent and does not apply on this record.

First, and unlike *Bartz*, direct evidence demonstrates that OpenAI could have used public domain works without impairing any transformative character of its LLMs. *See* SUF¶301-03. And *second*, evidence shows that OpenAI nonetheless trained on books because it reduced costs and made the model better at outputting substitutive material. *See supra* § I.A. But *Warhol*, 598 U.S. at 532, 547, is clear that enhancing commercially relevant product features is not enough to show justification. At best, the factual dispute regarding the connection of Plaintiffs' works to OpenAI's purportedly transformative ends must go to a jury. *McGucken*, 2022 WL 836786, at *13.

v. *Google Books* and the So-Called "Intermediate Copying" Cases Undermine, Rather Than Support, OpenAI's Fair Use Defense.

OpenAI also seeks to rely on cases where technology companies copied non-expressive functional elements from creative works, in service of a purpose entirely unrelated to the market for these works. Here, however, OpenAI copied the expressive elements of Plaintiffs' works to create a model to compete against those same works at the press of a button and whose outputs they sold to the public for fractions of a dollar. SUF¶845.

These cases embody a narrow principle: Where a party copies the functional elements of a work for a purpose entirely divorced from the work's protected, creative expression because the copying was necessary to enable a use, that may be transformative. As this court explained, cases involving exploitation of functional aspects of a work "must be interpreted in [their] particular context" as the "primarily functional" nature of the works "makes it difficult to apply traditional copyright concepts." *Prince*, 2023 WL 3383029, at *13 (discussing *Google v. Oracle,* 593 U.S. 1 (2021)).

On the other hand, where parties make "intermediate" copies for the purpose of exploiting creative content that copyright protects, absent a compelling justification, such copying is not fair use. This includes copies of scientific articles made by Texaco's scientists in the course of

conducting research, or course-packs of copyrighted material to be employed for classroom use, or a song used in a news broadcast. *Texaco*, 60 F.3d at 924; *Basic Books, Inc. v. Kinko's Graphics Corp.*, 758 F. Supp. 1522, 1531 (S.D.N.Y. 1991); *Byrne v. Brit. Broad. Corp.*, 132 F. Supp. 2d 229, 231-35 (S.D.N.Y. 2001) (Stein, J.).

The record is replete with internal documents, expert testimony, and fact witness testimony from which a reasonable juror could conclude that OpenAI derived value from the expressive elements of Plaintiffs' works *because* they enabled cheap training of LLMs optimized to generate competing works through a process of compression-memorization, and that any functional elements were copied only incidentally and were entirely unnecessary to any "generalized" or "general use" characteristics of the resulting LLMs. *See supra* § II.A.

Nor does *Google Books*—which "test[ed] the boundaries of fair use" 804 F.3d at 206—apply to OpenAI's conduct here. In *Google Books*—like the other intermediate copying cases—the copying in question was required to enable the use in question, the use in question relied upon the functional, as opposed to creative or expressive, aspects of the works, and there was no showing that the use had the possibility of displacing the market for the original. *Id.* at 206, 214-20, 227. In contrast to OpenAI's gratuitous copying, each copy Google made was functionally necessary to enable searching of *that particular book* and thus Google's overall transformative use.

OpenAI cites dicta in *Romanova*, 138 F.4th at 115-18, discussing *Google Books* to claim that its copying was justified because its models "render[] a valuable service to the public . . . without allowing public access to the copy," Mot. at 21, but that statement cannot be applied to the facts here. *Google Books* made clear that the fair use analysis as to storage of digital copies depended on the Plaintiffs' failure to show that Google insufficiently protected their works from exposure. 804 F.3d at 227-28. But, far from safeguarding Plaintiffs' works, OpenAI distributed

them to its business partners and third-party contractors, imposed no restrictions on internal access, and exposes training data to significant risks of exfiltration due to insufficient internal controls. RSUF¶34,104. And *Texaco*, 60 F.3d at 924, confirms that fair use does not attach simply because a company never exposed its internal "research" copies to the public.

Just as Google could not have a book search engine without copying books, the game developers in *Connectix*, 203 F.3d 596, and *Sega Enters. Ltd. v. Accolade, Inc.*, 977 F.2d 1510 (9th Cir. 1992), could not produce games without copying the functional code enabling interoperability; the search engine developers in *Perfect 10,* 508 F.3d 1146, and *Kelly v. Arriba*, 336 F.3d 811 (9th Cir. 2003), could not enable searching for a site's content without copying the content in question; and the Android API in *Oracle* could not call up the "implementing code" without copying the specific "declaring code," 593 U.S. at 11. The copying in each case served a practical, rather than "entertainment, aesthetic, or informative[,] function." *Perfect 10*, 508 F.3d at 1165.

The opposite is true here: The evidence above contradicts OpenAI's "functional" account of training and demonstrates that it exploited precisely the creative expressive content of Plaintiffs' works.

B. Factors Two and Three Favor Plaintiffs.

As to factor two and as set forth in Plaintiffs' opening Motion, Plaintiffs' books are at the core of copyright protection and OpenAI cannot seriously contend otherwise. OpenAI attempts to claim that it did not copy the protected expression for its expressive characteristics, but, as set forth above, that is entirely contradicted by the record.

As to factor three, OpenAI has two arguments: (1) that because it believes it needed "an enormous volume of text, use of entire books is reasonable," and (2) that it did not make the entirety of works available to the public. Mot. at 27–28.

As to (1), this is wrong for the reasons set forth above regarding OpenAI's (lack of) justification for its copying. *See supra* § II.A(c).

As to (2), while distribution of works to third parties certainly is an infringement—and one which OpenAI committed here by distributing the works to third-party contractors—the lack of such distribution does not imply that a use was fair. *See, e.g., Texaco*, 60 F.3d at 931–32.

C. Factor Four Weighs Strongly Against Fair Use

Factor Four of the fair use analysis requires examination of the "the effect of the use upon the potential market for or value of the copyrighted work." 17 U.S.C. § 107(4). The proponent of the fair use to must prove that their use "does not compete in the relevant market[s]," which includes "both the market for the original work *and* the market for any derivative works the rightsholder might develop." *Internet Archive*, 115 F.4th at 189 (cleaned up). This burden requires OpenAI to prove not only the lack of any effect of *its* use, but also that "unrestricted and widespread" conduct like OpenAI's would similarly cause no effect on Plaintiffs' original and derivative markets. *Id.*

"While the rightsholder may bear some initial burden of *identifying* relevant markets, we have *never held* that the rightsholder bears the burden of showing actual market harm." *Id.* at 194 (emphasis added; internal quotations omitted). Rather it is OpenAI's burden to disprove market impact. *See id.* at 193–95.

OpenAI falls far short of that burden as to both the harm to the market for licensing Plaintiffs' books for AI training and the harm to Plaintiffs' traditional books market due to the flood of AI-generated works.

OpenAI's makes two factual arguments as to the already-existing AI-training licensing market; both are contradicted by the record.

First, OpenAI argues that the licensing market does not exist. Mot. at 37–38. But significant evidence demonstrates that the market for licensing AI training data and, specifically, for licensing books as AI training data, not only exists but is rapidly growing, with more than [redacted] in book licensing agreements signed to date. SUF¶775–840. OpenAI and Microsoft themselves participate in this market, with Microsoft [redacted]. *See id.;* RSUF¶121. And OpenAI and Microsoft executives continually expressed interest in licensing, with one OpenAI employee testifying that OpenAI should have been willing to pay "10X of the cost of training a model" for a "high-quality, diverse e-books dataset." SUF¶297,775–814. Analysis from Plaintiffs' expert Michael Lasinski shows that even existing AI-training licenses cover the conduct Plaintiffs challenge here: the reproduction of their works to train LLMs capable of substituting for their works in the market for books. *See* Lasinski Decl. Ex. B ¶¶49–138. And this market possesses all the characteristics—widespread intermediaries capable of aggregating rights, independent collective licensing organizations, high-value materials, and clearly defined rights—necessary for success. ASUF¶52–56. Factual disputes thus preclude a finding that no licensing market exists.

Second, OpenAI argues that the licensing market is not "likely to be developed" because Plaintiffs have not participated in it. Mot. at 37. Contrary to OpenAI's suggestion, Plaintiffs repeatedly testified that they would be willing to license their works for AI training. *See* ASUF¶18–24. In any event, the Plaintiffs' decision to participate—or not—in the *existing* licensing market has no bearing on the Court's decision as that is one of the rights accorded to them by copyright. *See Castle Rock Ent., Inc. v. Carol Publ'g Grp., Inc*., 150 F.3d 132, 145–46 (2d Cir. 1998) (potential derivative market is not illusory just because an author "has evidenced little if any interest in exploiting this market for derivative works.").

OpenAI similarly cannot meet its burden as to Plaintiffs' original market. OpenAI principally relies on the expert testimony of Dr. Gregory Leonard to argue that the torrent of AI-generated works are not impacting the traditional market for books. Mot. at 34–35. Yet Leonard failed to even *analyze* the market for e-books or address uncontested [redacted] data demonstrating a flood of AI-generated works [redacted], *despite admitting that AI-generated books are likely digitally distributed*. RSUF¶¶125–131.

As *Internet Archive*, 115 F.4th at 191–92, makes clear, Leonard's failure to consider the ebook market and [redacted] data means that his opinions cannot satisfy OpenAI's burden to show a lack of harm as a matter of law.

Against this failure of proof, a stack of record evidence confirms that AI-generated works are harming the traditional market for books.

First, analysis by Mr. Lasinski correcting Dr. Leonard's omissions of the e-book market shows that Plaintiffs have experienced sales declines following the introduction of OpenAI's models. RSUF¶¶125–131. The [redacted] data neglected by OpenAI's experts bolsters this analysis [redacted]. SUF¶849–51.

Second, the numerous incidents of AI-generated works winning prizes, securing major publishing deals, and rising to new peaks of sales, similarly demonstrates that OpenAI's use is displacing the market for Plaintiffs' works and, in any event, affecting the "value" of Plaintiffs' works in that market. *See* SUF¶885–890; ASUF¶40–49.

Third, Dr. Shan's testing demonstrates that the in-suit models are capable of generating books which compare favorably with Plaintiffs' works all while mimicking Plaintiffs' expression

across a variety of genres, that this capability has increased every model generation (i.e. from GPT-3 to 3.5 to 4 to 4o), and will continue to advance. *See* Shan Decl. Ex. B. ¶10, § VI.

Fourth, OpenAI claims that Plaintiffs are not aware of lost sales from AI-generated books. Mot. at 35. But that is not Plaintiffs' burden and the Second Circuit has expressly rejected the proposition that Plaintiffs' must identify specific lost sales. *See Internet Archive*, 115 F.4th at 193. Regardless, the record flatly refutes OpenAI's claim. Plaintiff James Shapiro testified consistently with most other Plaintiffs: "AI will do and is doing harm to me as an author by making available my words and ideas and style and discoveries that are hard-earned and for which I paid, when necessary, to cover copyrighted materials. … I believe that it harms and will harm my ability to write and sell books." ASUF¶39; *see also* ASUF¶26–27 (Baldacci); ASUF¶34–35 (R.R. Martin); ASUF¶ 28 (Branch), ASUF¶29–30 (Day); ASUF¶31 (Golden); ASUF¶32 (Greer); ASUF¶33 (Grisham); ASUF¶36 (Schiff); ASUF¶37–38 (Connelly). A jury could easily credit the testimony from the Plaintiffs that AI generated books are substituting for and harming the market for their works (and the works of other authors), or that they were likely to do so in the future. *See* ASUF¶25–39;40–49.

Particularly in light of the statements from OpenAI and Microsoft executives recognizing the substitutive capacities of their models and targeting specific Plaintiffs for replacement, a factfinder could easily conclude that OpenAI's use of Plaintiffs' books is harming their original markets on any one of the above grounds. *See supra* § II.A(a)-(b). "Under these circumstances, whether and to what extent this factor favors a finding of fair use presents an issue for trial." *McGucken*, 2022 WL 836786, at *15.

i. OpenAI's Failure to Participate In The Existing and Growing AI-Training Data Licensing Market Is Dispositive On Factor Four.

Against the evidence above, OpenAI's responds that licensing market harms are "circular" and not cognizable because its use was transformative. *See* Mot. at 36–37. That is wrong.

Campbell, 510 U.S. at 590–93, and its progeny require consideration of existing or "likely to develop markets," even where the use may be transformative. *See id.*; *Texaco,* 60 F.3d at 930–31; *TVEyes,* 883 F.3d at 179–80. *Campbell*'s finding that the licensing market for parodies was "unlikely" to develop rested on the common-sense observation that rightsholders are unlikely to license works which have a "critical aspect." 510 U.S. at 592–93. But *Campbell* also declined to approve summary judgment of fair use because the defendant had not carried its burden as to harms to "protectible markets for derivative works," including the "derivative market [to license the work] for" non-parodic uses. *Id.* at 592–93.

Nothing about OpenAI's use implicates *Campbell's* criticism-focused distinction between "likely" and "unlikely" markets. Instead, as the record evidence cited above demonstrates, the AI-training data market is not only "likely" to exist, it, in fact, exists already. The existence and growth of that market suffices to distinguish the instant case from those like *White v. West Pub. Corp.*, 2014 WL 3385480, 3 (S.D.N.Y. July 11, 2014) and *Blanch v. Koons*, 467 F.3d 244, 258 (2d Cir. 2006) in which no likely market was found and thus no harm proven.

More recently and in the context of AI training, *Ross* held that because "the market for licensing headnotes as text to train AI is rapidly developing," harms to that market from the defendant's uncompensated use were cognizable and counseled against fair use even though the plaintiff had not licensed its works to others. *See* 2026 WL 2932557, at *9. The same is true on the record here. *See supra* §II(C).

As to *Google Books* and as noted above, the court grounded its holding regarding book-market impact on the fact that Google was using "unprotected fact[s]" about the books rather than,

as the record here shows, the copyrighted creative expression. 804 F.3d at 224. But *Google Books* did not hold harms to the licensing market were categorically non-cognizable, but instead that the "far more extensive" "paid licensing arrangements Plaintiffs cite" failed to demonstrate the existence of a market for Google's specific use. *Id.* at 226. Here the licenses proliferating across the AI industry provide similar scope to use books in AI training as OpenAI made here. SUF¶¶790–792,814,825.

And as to *Kadrey v. Meta,* 788 F. Supp. 3d 1034, and *Bartz,* 787 F. Supp. 3d 1007, and as explained above with respect to factor one, both cases rest on a materially different record regarding transformativeness and purpose lacking the evidence here showing that OpenAI's purpose in using the Plaintiffs' works was precisely to supplant and substitute for the works in question. *See supra* §II.A(a)-(c). Instead, and as *Ross* confirms, when evidence of substitutive purpose and an existing market are both present—as they are here—factor four weighs strongly in the Plaintiffs' favor. *See* 2026 WL 2932557, at *9.

ii. The Court Must Consider Harms From The Flood of AI-Generated Works Resulting From OpenAI's Use.

OpenAI contends that the Court should simply disregard the substantial evidence recounted above demonstrating the impact of AI-generated works on the original market for Plaintiffs' books because any harms stemming from non-infringing works are not cognizable. But this argument confuses the statutory requirements for establishing copyright *infringement*, 17 U.S.C. § 106, with questions of fair use, 17 U.S.C. § 107.

For a *prima facie* case, Plaintiffs must establish substantial similarity, which Plaintiffs' have easily done. *See* Pls. Mot. at 18-19. But that is separate from factor four, which asks about the "effect" of the use—here copying Plaintiffs' books to create a book-generating machine—on "potential market for or value of the copyrighted work." 17 U.S.C. § 107(4). The plain meaning

of "effect"—the "result, outcome, or consequence" of an action, "Effect," *Black's Law Dictionary* (12th Ed. 2024)— encompasses *all* substitution created by OpenAI's use, including harm from the flood of AI-generated books.

The Copyright Act, precedent and common-sense thus require the Court to reject Defendants' invitation to "stick [its] head[] in the sand" and consider the "obvious way that a new technology might severely harm the incentive to create." *Kadrey*, 788 F. Supp. 3d at 1055.

In *Ross*, 2026 WL 2932557 at *9, the court had no trouble finding that harms which stemmed from copying works "to create a competing legal-research platform" were cognizable and weighed in favor of the plaintiff even where the competing product did not serve the plaintiff's underlying copyrighted material to users. *Ross*, 2026 WL 2932557 at *9. Moreover, "unrestricted and widespread copying of [the plaintiff's] headnotes to develop substitute legal-research tools, like [the defendant] did, would harm [the plaintiff's] position in that original market." *Id.* The same is true here. The flood of AI-generated books into Plaintiffs' market, *see supra* § II.D, confirms that permitting any LLM company to train models on Plaintiffs' works would harm the "market for or value of" those works.17 U.S.C. § 107(4).

And in *Kadrey*, Judge Chhabria held unequivocally that factor four considers harms resulting from the sale of books "written by an LLM instead of … a human author" because such "indirect substitution is still substitution." 788 F. Supp. 3d at 1054. As *Kadrey* put it: "[n]o matter how transformative LLM training may be, it's hard to imagine that it can be fair use to use copyrighted books to develop a tool to make billions or trillions of dollars while enabling the creation of a potentially endless stream of competing works that could significantly harm the market for those books." *Id.* at 1059. The record above confirms that OpenAI's LLMs have the ability "to flood the market with endless amounts . . . [of] books" which "compete for sales with

the books in the training data" and thus "reduce the incentive for authors to create—the harm that copyright aims to prevent." *Id.* at 1034, 1054. The fourth factor thus weighs strongly against OpenAI's use of Plaintiffs' works to train the in-suit LLMs. *Id.* at 1034, 1054.

While OpenAI relies upon *Sega v. Accolade, Inc.*, *Sega* involved an attempt by the plaintiffs to bootstrap copyright into a monopoly over the functional or non-expressive elements of computer code. *See* 977 F.2d at 1527. *Sega* held only that factor four favored copying of non-expressive functional code enabling operation of game cartridges *because* (1) there was "nothing in the record that suggest[ed] that [the defendant] had copied any" of the works' creative elements and (2) there was no evidence the copying had "significantly affected the market for Sega's [games]." *Id.* at 1523.

Here, however, the record shows OpenAI copied Plaintiffs' creative, expressive content to train a machine to displace Plaintiffs' works and the record contains substantial evidence that AI-generated books created by OpenAI's LLMs are flooding the market and substituting Plaintiffs' works. *See supra* § II.A(a).

The same is true of *Google Books*, 804 F.3d 202. The court based its factor four search-and-snippet holding on the fact that Google's use did not "serve as a substitute for the books that are being searched" because (1) it would cost too much time and effort to obtain substitutive output, (2) as a result any lost sales would be *de minimus* and, (3) any harm would result from Google's use and exposure of unprotected facts. *Id.* at 223-25. All three prerequisites are missing here. OpenAI's models *can* produce competitive outputs in minutes and for pennies. *See supra* § II.A(a). AI-generated works *are* having a seismic impact on the market for books. *See supra id*. And the evidence recounted above demonstrates that these competing outputs derive *directly* from

the exploitation of Plaintiffs' protected expression during training rather than the indexing of unprotected facts. *See supra* § II.A(a)-(c). *Google Books* simply does not apply to this record.

As to *Bartz*'s contrary treatment of market harm, 787 F. Supp. 3d at 1031-32, and while much of Judge Alsup's opinion is consistent with longstanding principles of fair use, its sole citation for its market dilution holding was *Sega*, distinguished above. *Bartz*'s market dilution holding rests on the same confusion between infringement and fair use detailed above. *See id.* at 1021 ("Authors challenge only the inputs, not the outputs, of these LLMs"). To the extent *Bartz* can be read to suggest that AI-generated displacement is categorically non-cognizable, that error rests on a flawed premise. OpenAI's LLMs are machines capable of generating an infinite amount of book content, with minimal effort and at minimal cost; they are not children learning how to read and write. *See supra* § II.A(a)-(b).

OpenAI's other citations are similarly unavailing. *Sony Comput. Ent. Am., Inc. v. Bleem, LLC*, 214 F.3d 1022, 1029 (9th Cir. 2000), and *Triangle Publ'ns, Inc. v. Knight-Ridder Newspapers, Inc.*, 626 F.2d 1171, 1177 (5th Cir. 1980), concern alleged infringement by advertisements and so *Campbell's* focus on protecting criticisms and comparisons foreclosed the harms. Similarly, in *NXIVM Corp. v. Ross Institute*, 364 F.3d 471, 482 (2d Cir. 2004) and *Bill Graham Archives,* 448 F.3d 605, the uses—criticism by a cult blogger and work of history, respectively—fell squarely within *Campbell's* protection and "require[d] incorporation [of the works] . . . for optimum treatments of their subjects." *See Prince*, 2023 WL 3383029, at *21 (cleaned up). Here, and as explained above, *supra* § II.A(a)-(c), nothing about OpenAI's use "require[d] incorporation" of Plaintiffs' works. *Id.*

Finally, OpenAI suggests that Plaintiffs' position would "prohibit[] competition from new expression." Mot. at 33. Precisely the opposite. OpenAI's position—that it can copy any work it

chooses in the course of building a machine capable of displacing the market for the very expression copied and solely on the grounds that doing so is easier, cheaper, and more profitable than paying for a license—threatens, if "unrestricted and widespread"—to destroy all returns to human authorship. In these circumstances, the court must heed the statutory and precedential command to evaluate all effects of OpenAI's use on the market for and value of Plaintiffs' works.

D. Public Benefits

In light of the above—and that any benefits from OpenAI's LLMs can be obtained without exploiting Plaintiffs' expressive content—the Court need not credit or consider OpenAI's claimed benefits. *See* Mot. at 38-39. However, if considered, Defendants' own executives have publicly described LLM technology as a potentially existential threat to human beings and research in the field has documented harms to mental health, to the labor market, and from increased cyberattack and pandemic risks as a result of Defendants' conduct. *See* SUF¶898-916; ASUF¶62-64.

III. Use #3: OpenAI's Motion Entirely Fails to Address Its Distribution.

As set forth in Plaintiffs' Opening Motion and Plaintiffs' Opposition to Microsoft's Motion (incorporated by reference), OpenAI's distributions of Plaintiffs' works for commercial ends unconnected to any training purpose is a separate—and not fair—use. *See* Pls. Mot. at 38-39.

Yet OpenAI's only argument as to distribution appears in a footnote asserting—without citation—that any distribution served the "same transformative purpose" as training and that Plaintiffs did not plead such a claim. Mot. at 24 n.15. But as the Court made clear when rejecting OpenAI's last motion to strike, Plaintiffs pled a direct copyright infringement claim under 17 U.S.C. § 106 and there was no need to specify the precise legal theories of relief. *See* Dkt. 707 at 4-5. By relegating this issue to a footnote, OpenAI has waived any argument, *see, e.g., Putnam Advisory*, 2020 WL 264146, at *2, and the Court can deny OpenAI's motion on that basis.

CONCLUSION

The Court should deny OpenAI's motion for summary judgment and grant Plaintiffs' motion for partial summary judgment.

Dated: October 29, 2026

/s/ Justin A. Nelson
Justin A. Nelson
SUSMAN GODFREY L.L.P.
1000 Louisiana Street, Suite 5100
Houston, TX 77002
Telephone: 713.651.9366
jnelson@susmangodfrey.com
Interim Lead Counsel for Class Plaintiffs

CERTIFICATE OF WORD COUNT

In accordance with Local Civil Rule 7.1(c) and the Court's Order at Dkt. 1664, 25-md-3143, I certify that the foregoing Memorandum of Law is 12,498 words, including footnotes but exclusive of the caption page, table of contents, table of authorities, certificates and signature block. The basis of my knowledge is the word count feature of the word-processing system used to prepare this memorandum.

*/s/J. Craig Smyser*__
J. Craig Smyser

CERTIFICATE OF SERVICE

I hereby certify that on October 29, 2026, I caused the foregoing document to be electronically filed with the Clerk of the United States District Court for the Southern District of New York using the CM/ECF system, which shall send electronic notification to all counsel of record.

*/s/J. Craig Smyser*__
J. Craig Smyser