UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

IN RE: OPENAI, INC., COPYRIGHT INFRINGEMENT LITIGATION This Document Relates To: 1:23-cv-08292-SHS-OTW 1:23-cv-10211-SHS-OTW 1:25-cv-03482-SHS-OTW	25-md-3143 (SHS) (OTW) Hon. Sidney H. Stein Hon. Ona T. Wang ORAL ARGUMENT REQUESTED PUBLIC VERSION - REDACTED

OPENAI'S OPPOSITION TO CLASS PLAINTIFFS'

MOTION FOR SUMMARY JUDGMENT

TABLE OF CONTENTS

I. INTRODUCTION 1

II. BACKGROUND 3

A. OpenAI's Training Process 3

B. OpenAI Uses Multiple Techniques to Prevent Regurgitation, and Regurgitation Is Rare 4

C. Books Were Only One of Many Sources Used to Train OpenAI's LLMs 5

D. Plaintiffs' Claims of Harm 6

E. Each Alleged Distribution of Books Was for the Purpose of Training LLMs 6

III. ARGUMENT 8

A. The Proper Scope of the Fair Use Inquiry Requires Analyzing the Ultimate Purpose of the Use, Not Each Act of Copying in Isolation 8

1. The Fair Use Analysis Must Be Organized Around the Purpose of the Use, Not Each Act of Copying in Isolation 9

2. Plaintiffs' Authorities Do Not Support a Copy-by-Copy Approach 10

B. OpenAI's Alleged Use Was Fair Use 15

1. Factor One: OpenAI's Use of the Asserted Works Was Highly Transformative 15

a. Plaintiffs' Argument that OpenAI Designed its Products to be Capable of Generating Competing Outputs is Incorrect and Irrelevant 16

b. Plaintiffs Misinterpret the "Reasonable Necessity" Inquiry 21

c. Commerciality Does Not Tip Factor One Against Fair Use 22

d. Plaintiffs' Assertions Regarding "Piracy" Do Not Change the Fair Use Analysis 23

2. Factor Two: OpenAI's Alleged Copying Was Not to Capture or Republish Protected Expression 24

3. Factor Three: OpenAI Used a Reasonable Amount of the Asserted Works in Light of Its Transformative Purpose 25

4. Factor Four: OpenAI's Alleged Use Did Not Cause Cognizable Harm 26

a. "Market Dilution" is Not a Cognizable Theory of Harm 26

b. Plaintiffs Have Not Suffered *Any* Harm, Cognizable or Not....30

c. Plaintiffs Have Not Suffered Harm to Any Cognizable Licensing Market 33

d. Alleged Lost Sales From the Uses at Issue are Not Cognizable and are Insufficient as a Matter of Law 36

C. Plaintiffs Cannot Assert an Unpleaded Distribution Claim....37

D. Plaintiffs Cannot Rely on Models Outside the Scope of this Case....37

IV. CONCLUSION....38

TABLE OF AUTHORITIES

Page(s)

Federal Cases

A&M Recs., Inc. v. Napster, Inc.,
239 F.3d 1004 (9th Cir. 2001) 12

Am. Geophysical Union v. Texaco Inc.,
60 F.3d 913 (2d Cir. 1994) 11, 12, 34

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

Arica Inst., Inc. v. Palmer,
970 F.2d 1067 (2d Cir. 1992) ..... *passim*

Authors Guild, Inc. v. HathiTrust,
755 F.3d 87 (2d Cir. 2014) 10, 33, 35

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

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) 11

Bill Graham Archives v. Dorling Kindersley Ltd.,
448 F.3d 605 (2d Cir. 2006) 25, 33, 34, 36

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

Fox News Network, LLC v. TVEyes, Inc.,
883 F.3d 169 (2d Cir. 2018) 11

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

Hachette Book Grp., Inc. v. Internet Archive,
115 F.4th 163 (2d Cir. 2024) 12, 31, 34, 37

Harper & Row Publishers, Inc. v. Nation Enters,
471 U.S. 539 (1985) 37

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

Kane v. Comedy Partners,
2003 WL 22383387 (S.D.N.Y. Oct. 16, 2003) 20

Lennon v. Premise Media Corp.,
556 F.Supp.2d 310 (S.D.N.Y. 2008) 24

Monster Commc'ns, Inc. v. Turner Broad. Sys., Inc.,
935 F.Supp. 490 (S.D.N.Y. 1996) 28

NXIVM Corp. v. Ross Inst.,
364 F.3d 471 (2d Cir. 2004) 14, 23, 24, 29

Perfect 10, Inc. v. Amazon.com, Inc.,
508 F.3d 1146 (9th Cir. 2007) 10, 14, 23

Romanova v. Amilus Inc.,
138 F.4th 104 (2d Cir. 2025) 22

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

Estate of Smith v. Cash Money Records, Inc.,
253 F.Supp.3d 737 (S.D.N.Y. 2017) 20

Sony Comput. Ent. Am., Inc. v. Bleem, LLC,
214 F.3d 1022 (9th Cir. 2000) 28

Sony Computer Ent., Inc. v. Connectix Corp.,
203 F.3d 596 (9th Cir. 2000) 10, 26, 28

Thomson Reuters Enter. Centre GmbH v. ROSS Intel. Inc.,
F.4th -, 2026 WL 2932557 (3d Cir. Sept. 29, 2026) 20, 34

Triangle Publ'ns, Inc. v. Knight-Ridder Newspapers, Inc.,
626 F.2d 1171 (5th Cir. 1980) 28

UMG Recordings, Inc. v. MP3.Com, Inc.,
92 F. Supp. 2d 349 (S.D.N.Y. 2000) 34

White v. West Pub. Corp.,
2014 WL 3385480 (S.D.N.Y. July 11, 2014) 34

Wright v. Ernst & Young LLP,
152 F.3d 169 (2d Cir. 1998) 37

Wright v. Warner Books, Inc.,
953 F.2d 731 (2d Cir. 1991).....27, 29

Federal Statutes

17 U.S.C. § 106.....37

17 U.S.C. § 107..... *passim*

Rules

Fed. R. Evid. 702 6

Other Authorities

Brief of Plaintiffs Appellant*s, Authors Guild v. Google, Inc.*,
2014 WL 1509692 (2d Cir., April 11, 2014) 9

Decl. of Dan Clancy ¶ 12, *Authors Guild v. Google*,
No. 05-cv-08136, ECF 1035.....14

Brief of Plaintiffs-Appellee*s, Sony Comput. Ent. Inc. v. Connectix Corp.*,
1999 WL 33623858 (9th Cir. June 24, 1999) 23

Pierre N. Leval, *Campbell as Fair Use Blueprint?*, 90 Wash. L. Rev. 597, 612–13
(2015).....24

I. INTRODUCTION[1]

The parties agree on one thing: the question of fair use should be resolved on summary judgment. But Plaintiffs get the operative legal doctrine wrong. Three key principles, long-established in binding precedent, dictate that OpenAI's use was fair use as a matter of law:

First, the fair use analysis is holistic, not atomistic. When a defendant is accused of infringing copyright as part of the internal process of developing a new technology, courts assess the lawfulness of those internal copies in light of the purpose of the use—not, as Plaintiffs suggest, by artificially truncating the analysis to begin and end with the observation that internal copies were made, without regard to why and to what end. As one court explained in the context of another case over LLM training, "[t]o say that [the defendant's] downloading was 'piracy' and thus cannot be fair use begs the question because the whole point of fair use analysis is to determine whether a given act of copying was unlawful." *Kadrey v. Meta Platforms, Inc.*, 788 F.Supp.3d 1026, 1046 (N.D. Cal. 2025).

Second, using copyrighted content for the purpose of creating a new technology like an LLM is highly transformative and thus favors fair use. Courts have long recognized that copying works internally to develop technology serving a fundamentally different purpose favors fair use, particularly where the resulting technology does not materially expose the original expression to users. This precedent runs from *Sega v. Accolade* in the Ninth Circuit (1992) through *Google Books* in the Second (2015), and most recently, two decisions in 2025 specifically addressing the use of books to train LLMs—which the courts found "highly," "quintessentially," and "spectacularly" transformative. *Kadrey*, 788 F.Supp.3d at 1044; *Bartz v. Anthropic PBC*, 787 F.Supp.3d 1007, 1021–22, 1026 (N.D. Cal. 2025).

[1] In all quotations appearing herein, all internal quotation marks, citations, alterations, and the like have been omitted, and all emphases added, unless otherwise noted.

Third, Plaintiffs have not suffered any cognizable fourth factor harm. Plaintiffs do not claim lost sales, declining prices, diminished market share, or any tangible harm from OpenAI's alleged use of their books for pretraining. They concede they have no "empirical evidence" of harm. Instead, they invoke purported "common sense" to speculate that AI-generated books may compete with their own. But competition from new, non-infringing works (even if established) is not the type of market harm copyright protects against. Nor can Plaintiffs establish harm by pointing to license fees OpenAI did not pay for its transformative use. This argument assumes its conclusion: that OpenAI's use required a license in the first place. Indeed, both *Bartz* and *Kadrey* rejected similar claims of harm based on alleged lost licensing fees for LLM training.

Plaintiffs' contrary view of the law would improperly expand copyright into a broad *de facto* veto over new LLM technologies. On Plaintiffs' view, using books to create any general-purpose technology capable of writing is not transformative if that technology could be used to produce original, non-infringing, competing books. And these new books, created by the technology's users rather than OpenAI, would somehow also constitute a fourth factor harm by diluting the market for existing books. If accepted, these propositions together stand for the principle that using books to create a general-purpose technology capable of writing (irrespective of its other capabilities) can never be a fair use. The exclusive rights of book authors surely do not extend so far.

The ultimate question of fair use is for the Court. *See Google LLC v. Oracle Am., Inc.*, 593 U.S. 1, 23–24 (2021). And on a sound application of settled law, OpenAI's alleged use of Plaintiffs' copyrighted works to train its LLMs was fair. For these reasons, and the reasons below, Plaintiffs' motion should be denied, and OpenAI's motion should be granted.

II. BACKGROUND

OpenAI's summary judgment motion provides a comprehensive background of the relevant facts. OpenAI incorporates by reference its summary judgment papers, including its Statement of Undisputed Material Facts. *See* MDL ECF 1718 ("OpenAI-Class Mot.") at 2–16; MDL ECF 1735 ("OSUF").

A. OpenAI's Training Process

Plaintiffs' motion largely ignores OpenAI's LLM training process. It is undisputed that to train the In-Scope Models,[2] OpenAI used enormous datasets, comprising up to tens of trillions of words drawn from numerous, diverse sources across the Internet. OSUF ¶¶ 23–30.

OpenAI then de-duplicated, processed, and tokenized the data before splitting it into excerpts used for pretraining. OSUF ¶¶ 23, 30–38. During pretraining, the model used each excerpt to generate a probability distribution predicting the next token. The training algorithm then compared that prediction against the actual next token, measured the difference, and used that result to incrementally update the model's parameters. This process repeated millions of times across many different excerpts. MDL ECF 1848 ("Class Mot.") at 10 (citing MDL ECF 1988 ("CSUF") ¶¶ 95–100); OpenAI-Class Mot. at 6 (citing OSUF ¶¶ 39–44).

OpenAI then evaluated the model's performance using a separate "test set" of data never shown to the model before. The resulting metric, called "test" or "validation" loss, measured the model's ability to generalize—to learn patterns that can be applied to new, unseen data. This is a core measure of pretraining performance. OSUF ¶¶ 45–48.

[2] The In-Scope Models are GPT-3, GPT-3.5, GPT-3.5 Turbo, GPT-4, GPT-4 Turbo, GPT-4o, and GPT-4o mini. MDL ECF 707 at 1. Plaintiffs assert, however, that "GPT-4o mini did not train on any data containing the Asserted Works." CSUF ¶ 534. The Court should therefore grant summary judgment of non-infringement as to that model.

OpenAI also developed numerous innovative post-training techniques that teach the model to engage in a question-and-answer dialogue and function as a useful assistant. OSUF ¶ 12. And OpenAI conducts safety evaluations, including human "red team[]" testing. OSUF ¶¶ 51–52.

Together these processes serve a fundamentally different purpose from the original works: developing general-purpose software capable of understanding and generating natural human language.

B. OpenAI Uses Multiple Techniques to Prevent Regurgitation, and Regurgitation Is Rare

OpenAI has developed multiple techniques to mitigate potential "regurgitation," including de-duplicating training data, modifying the training process to emphasize generalizable patterns, post-training models to refuse requests for copyrighted text, and implementing an output filter ("Bloom filter") to prevent training data from appearing in outputs. OSUF ¶¶ 58–81.

The real-world evidence confirms the effectiveness of these measures. OpenAI's models rarely generate text that matches text from Plaintiffs' works, and when they do, the excerpts are exceedingly short. *See* OSUF ¶¶ 81–88. OpenAI's expert examined approximately 95 million outputs from 20 million ChatGPT conversations and found only 14 instances containing verbatim or near-verbatim regurgitation from the Asserted Works—*i.e.*, an alleged per-conversation regurgitation rate of 0.00007%. All but two were 32 words or fewer, and none amounted to more than 0.017% of an Asserted Work. OSUF ¶¶ 82–84. Plaintiffs' own expert examined approximately 338 million outputs from 108 million conversation logs and identified reproductions of 30 or more words from just seven of Plaintiffs' works in only 31

conversations—an even lower per-conversation regurgitation rate of 0.00003%. OSUF ¶ 91; RCSUF[3] ¶ 986.

Even when Plaintiffs' expert deliberately attempted to elicit regurgitation through more than 5.3 million prompts, the expert generated only three excerpts that exceeded 300 contiguous words. OSUF ¶ 85; MDL ECF 1750 ("Berg-Kirkpatrick Decl.") ¶ 70. The longest contiguous passage generated was 1,899 words from *A Game of Thrones*—just 0.62% of the work. OSUF ¶¶ 86–88. Each of the three excerpts came from a George R.R. Martin book, was fully available through ordinary internet searches or publicly accessible websites (including Amazon Kindle samples), and accounted for less than 1% of the content of the corresponding Asserted Work. Berg-Kirkpatrick Decl. ¶¶ 70–71, Table 4; OSUF ¶ 89.

C. Books Were Only One of Many Sources Used to Train OpenAI's LLMs

Plaintiffs assert that books are uniquely "high-quality" training data. Class Mot. at 11. But OpenAI's research has shown that size and diversity of the overall pre-training corpus, rather than any one particular work, affects model performance. OpenAI-Class Mot. at 3, 5; OSUF ¶¶ 9–10; RCSUF ¶¶ 942–944 (explaining that OpenAI sought to obtain "a diverse set of text over many topics"; books were simply another "natural topic of diverse material" to include alongside data such as source code, math problems, legislative materials, and public web data).

The record also shows that Plaintiffs overstate the extent to which books from particular sources were used to train the In-Scope Models. Among the In-Scope Models, only GPT-3 and GPT-3.5 were trained on material downloaded from LibGen. OSUF ¶ 11; *see also* CSUF ¶ 519. It is undisputed that OpenAI stopped using LibGen books to train models in late 2021, after

[3] Defendants' Response to Class Plaintiffs' Rule 56.1 Statement ("RCSUF").

GPT-3.5. CSUF ¶ 265. It is also undisputed that OpenAI did not use any books from Books3 to train any In-Scope Model. RCSUF ¶ 948.

D. Plaintiffs' Claims of Harm

OpenAI's economic expert, Dr. Gregory Leonard, analyzed the relationship between ChatGPT (and OpenAI's API) and sales and pricing trends for the Asserted Works specifically and books more generally, and found that OpenAI's products were not associated with a systematic decrease in sales trends and did not have a statistically significant impact on price trends. OSUF ¶¶ 125–127. He also conducted a control-group study comparing the Asserted Works against works not used for training; the evidence showed no statistically significant difference in sales. OSUF ¶¶ 128–129. Plaintiffs do not challenge Dr. Leonard's analysis in their motion, nor did they move to exclude it under Federal Rule of Evidence 702. MDL ECF 2609. Meanwhile, Plaintiffs' expert conducted no empirical lost-sales analysis of his own and testified that he could not identify any ChatGPT-prompted book that would have caused another author to lose a sale. OSUF ¶¶ 130–131. And Plaintiffs themselves could not identify lost sales attributable to ChatGPT. OSUF ¶¶ 113–119.

E. Each Alleged Distribution of Books Was for the Purpose of Training LLMs

Plaintiffs attempt to recast several transfers of training data as independent acts of "distribution," divorced from the purpose those transfers served. *See* Class Mot. at 38–39. But Plaintiffs concede that each transfer occurred in connection with training LLMs. *See*, *e.g.*, CSUF ¶¶ 573, 579, 589, 596, 601, 607. Each transfer was for the purpose of training, as explained below.

Project Mango. Plaintiffs concede that the project in which Microsoft and OpenAI crawled different parts of the internet ("Project Mango") was undertaken to "collect data for the purposes of OpenAI to train LLM models." CSUF ¶¶ 597–604.

GPT-3 Training Data. Plaintiffs allege that OpenAI provided Microsoft access to GPT-3's training data but concede the purpose was "so that Microsoft could '[redacted] including '[redacted]." CSUF ¶ 573 (alterations in original). OpenAI confirmed that the data transferred was only a ***subset*** of GPT-3's training data. RCSUF ¶¶ 573, 949. [4] Thus, the purpose of the transfer was to enable Microsoft to [redacted]. The later transfer of GPT-3's training data [redacted] served the same training purpose. *See* CSUF ¶ 579; RCSUF ¶¶ 579, 587, 950. The purpose of this [redacted] was to use the data to train models. RCSUF ¶ 951. Notably, this transfer excluded the LibGen-derived Books1 and Books2 datasets. *See* RCSUF ¶ 952.

Project Taxi. With respect to the project that used portions of Bing Index data ("Project Taxi"), Plaintiffs concede that its purpose was to assess data for training out-of-scope models. CSUF ¶ 596; *see also* RCSUF ¶¶ 593, 597, 953. OpenAI used the data to train "internal-only research models for evaluation purposes[,]" to determine whether it should be used to train OpenAI's "production models." RCSUF ¶ 954. OpenAI did not include the portions of Bing Index data in production-model training. RCSUF ¶¶ 597, 954–956.

Book-Summarization Project. Plaintiffs mischaracterize OpenAI's experimental "book-summarization project" as "distribut[ing] books" to third-party contractors. CSUF ¶¶ 607–608; Class Mot. at 10. OpenAI did not "distribute" books to contractors. Instead, OpenAI used a custom-built web portal with restricted login-only access to display very short snippets to contractors evaluating model-generated summaries. RCSUF ¶ 957. This research project tested

[4] Plaintiffs also concede that OpenAI's transfer of [redacted] of GPT-3's training data for "[redacted]" did not include books from LibGen. *See* RCSUF ¶¶ 587, 952.

novel, experimental training techniques—not used to train any In-Scope Model—by examining whether breaking difficult tasks into smaller, more readily evaluated subtasks could improve model training. *See* RCSUF ¶¶ 957–958.

III. ARGUMENT

OpenAI's alleged use of the Asserted Works was fair use as a matter of law.[5]

A. The Proper Scope of the Fair Use Inquiry Requires Analyzing the Ultimate Purpose of the Use, Not Each Act of Copying in Isolation

Plaintiffs approach the fair use inquiry through a faulty analytical framework in which every act of alleged copying—downloading, ingesting into a training pipeline, transferring between affiliated entities—is evaluated atomistically, rather than as an integrated course of conduct to be assessed as a whole. *See* Class Mot. at 18–19, 25, 38. Virtually all fair use decisions involving digital technology concern uses of copyrighted content whereby multiple internal copies are made in the service of offering a consumer-facing tool. Courts assessing the lawfulness of those multiple internal copies have not segregated the analysis to separately evaluate, *e.g.*, whether the initial copy acquired was fair use, a subsequent copy transferred to another server was fair use, and so on. Instead, courts assess whether all copying that was done for the "purpose" of creating the end product was fair use or not, in one overarching analysis.

Here, it is undisputed that OpenAI downloaded and made copies of copyrighted works for the sole purpose of training LLMs. Indeed, Plaintiffs concede that the sole purpose for which OpenAI allegedly downloaded books from LibGen was to train LLMs. *See* CSUF ¶ 186; *see also* CSUF ¶¶ 188, 195–200, 228. Plaintiffs make the same concession for books allegedly

[5] For purposes of this motion, OpenAI does not contest the elements of *prima facie* infringement as to the 194 Asserted Works the Class Plaintiffs have sought summary judgment on. As explained above, OpenAI's alleged use of those works is fair use as a matter of law. OpenAI reserves all defenses with respect to other works not at issue in Plaintiffs' motion.

downloaded from Books3. *See* CSUF ¶¶ 412, 402, 532. And unrebutted testimony from OpenAI confirmed that "the intention of all of the data that we collect is for the purposes of training." RCSUF ¶ 941. The other alleged acts of copying were similarly directed to training LLMs. *See supra* § II.E. The lawfulness of all these alleged copies must be assessed by the legality of that overarching training purpose, which was highly transformative.

1. The Fair Use Analysis Must Be Organized Around the Purpose of the Use, Not Each Act of Copying in Isolation

Section 107 is clear that the fair use analysis requires consideration of "the ***purpose*** and character of the use." 17 U.S.C. § 107(1). In keeping with that statutory mandate, binding authority dictates that fair use should be analyzed at the level of the ***purpose*** of the use, not at the level of each individual copy made to effectuate that use.[6] In *Campbell*, for instance, the Supreme Court assessed 2 Live Crew's overarching use of Roy Orbison's song as a parody—not as separate "uses" for the studio recording, the mixing, and the distribution of the final track. *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 583–89 (1994).

The Second Circuit followed that approach in *Google Books*. There, the plaintiffs argued that Google's mass scanning and storage of books on its servers was merely a "mechanical conversion" that could not be considered "transformative." Brief of Plaintiffs-Appellants, *Authors Guild v. Google, Inc.*, 2014 WL 1509692, at *32 (2d Cir., April 11, 2014). The Second Circuit, however, did not address the act of scanning in isolation and did not atomize Google's conduct into separate "uses" for copying scanned books to its servers, indexing them, or caching them. Rather, the court divided its analysis based on the ultimate purpose of each use—it

[6] This is why OpenAI's summary judgment motion on the News Plaintiffs' claims addresses pretraining and Browse separately. The purposes of those two uses are different, though both are highly transformative. *See, e.g.*, MDL ECF 1749 ("OpenAI-News Mot.") at 7.

separately assessed whether the "search function" and the "snippet view" were fair uses (concluding that they each were), rather than separately addressing each act of copying done to build those functionalities. *Authors Guild v. Google, Inc.* ("*Google Books*"), 804 F.3d 202, 217–18 (2d Cir. 2015); *accord Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87, 97 (2d Cir. 2014).

Out-of-circuit authorities also analyze fair use by looking to the purpose of the use rather than individual copies. In *Kadrey*, the court expressly rejected the argument that Meta's alleged downloading of books should be considered separately from its use of those books for LLM training. *Kadrey*, 788 F.Supp.3d at 1047. The *Kadrey* court held that the act of downloading "must still be considered in light of its ultimate, highly transformative purpose: . . . training [the model]." *Id.* This approach is consistent with well-established fair use opinions in the Ninth Circuit. *See, e.g.*, *Perfect 10, Inc. v. Amazon.com, Inc.*, 508 F.3d 1146, 1165 (9th Cir. 2007) (analyzing Google's creation of thumbnail images not as a series of separate technological copying events, but as a single transformative use for the purpose of providing a visual search engine); *Sony Computer Ent., Inc. v. Connectix Corp.*, 203 F.3d 596, 602–08 (9th Cir. 2000) (assessing multiple instances of copying in light of the overall purpose of the copying); *Sega Enters. Ltd. v. Accolade, Inc.*, 977 F.2d 1510, 1522–28 (9th Cir. 1992) (similar). Plaintiffs cannot defeat fair use by isolating the technical steps necessary to accomplish a transformative use.

2. Plaintiffs' Authorities Do Not Support a Copy-by-Copy Approach

Plaintiffs misread *Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith*, 598 U.S. 508 (2023), as requiring "each instance of copying" to be "analyzed independently." Class Mot. at 2, 19, 21. *Warhol* requires courts to distinguish different uses, not separately scrutinize every act of copying necessary to accomplish a particular use. The "use" at issue in *Warhol* was the licensing of "Orange Prince," a silkscreen made by Andy Warhol from a photograph taken by Lynn Goldsmith, to Condé Nast. *Warhol*, 598 U.S. at 526. The Court analyzed the overall

purpose of that use; it did ***not*** analyze separately each instance of copying that occurred in the course of that "use," such as the digitization, storage, and transmission of the work. *See id.*

Indeed, *Warhol* confirms that the relevant distinction is between uses serving different purposes. The Court explained that "[t]he same copying may be fair when used for one purpose but not another." 598 U.S. at 533. The Court therefore distinguished licensing Orange Prince to illustrate a magazine story about Prince (the use at issue) from other potential uses of the same work, such as creating the silkscreen in the first instance, displaying it in galleries or museums, using that work "for teaching purposes," or illustrating a magazine about Warhol himself (rather than Prince), *id.* at 534 & n.10, 536 n.12. Those are distinct ***uses*** because they serve different ***purposes***—not because they involve different acts of copying. Nothing in *Warhol* requires courts to subject each copy made in furtherance of the same single use to a separate fair-use analysis.

Plaintiffs likewise argue that a "downstream use" cannot "bless other acts of copying." Class Mot. at 21, 39. But none of Plaintiffs' cited cases supports their argument that the fair use analysis begins and ends with the bare fact that a copy was made. Indeed, the courts' reasoning in those cases had nothing to do with whether these purposes were "downstream" from the act of copying. Instead, the courts analyzed the purpose of the use to determine whether it is transformative. *See Am. Geophysical Union v. Texaco Inc.*, 60 F.3d 913, 924 (2d Cir. 1994) (finding a use not transformative where photocopies of articles were distributed to be read); *Basic Books, Inc. v. Kinko's Graphics Corp.*, 758 F.Supp. 1522, 1531 (S.D.N.Y. 1991) (finding a use not transformative where photocopied excerpts of textbooks were bound and sold to students to read); *see also Fox News Network, LLC v. TVEyes, Inc.*, 883 F.3d 169, 178 & n.4, 180–81 (2d Cir. 2018) (finding a use only modestly transformative where recordings of broadcast television were given to users in snippets to watch).

Texaco illustrates the point. It stands for the basic proposition that the court should focus on the "purpose . . . of the use," not the defendant's line of business or general occupation. *Texaco*, 60 F.3d at 918 (quoting 17 U.S.C. § 107(1)). The defendant in *Texaco* made copies of the works so that its employees could have them "readily available in their own offices" without purchasing additional subscriptions. *Id.* at 918–19. That was not a transformative use. *Id.* at 923. And the fact that the employees who received the copies happened to be researchers did not change the result. *Id.* at 924. But OpenAI is not claiming that its copying is fair because the copies were made by people engaged in research. OpenAI's use is fair because it made alleged copies of Plaintiffs' books for the ***purpose*** of training, *i.e.*, extracting uncopyrightable information from a larger corpus of LLM training data—not to supply the books for reading (as the Texaco employees did). The purpose of OpenAI's use was fundamentally different from the purpose of the original works.

Nor do *Napster* or *Hachette* support Plaintiffs' "download-only" theory. Both cases focused on quintessential instances of sharing the copyrighted expression itself with end users: music to listen to and books to read. *See A&M Recs., Inc. v. Napster, Inc.*, 239 F.3d 1004, 1015 (9th Cir. 2001), as amended (Apr. 3, 2001) (copies of music were downloaded and distributed for users to listen to); *Hachette Book Grp., Inc. v. Internet Archive*, 115 F.4th 163, 181 (2d Cir. 2024) ("IA's digital books serve the same exact purpose as the originals: making authors' works available to read."). In contrast, and as explained below, OpenAI's alleged use is transformative—OpenAI used the Asserted Works not to be read as written, but to train LLMs. Those LLMs are designed to produce original works, not regurgitate protected expression to users.

Plaintiffs' heavy reliance on *Bartz* fares no better. Class Mot. at 21–22. There, Anthropic downloaded books from Libgen to build a permanent "central library" for "any number of further uses" including purposes unrelated to model training. *Bartz*, 787 F.Supp.3d at 1014–16, 1020, 1026–27. The Court distinguished that library-building purpose from training LLMs, which it treated as a separate "further use." *Id.* at 1026 ("The objective analysis here shows the initial copies were pirated to create a central, general-purpose library, as a substitute for paid copies to do the same thing."). That distinction supports OpenAI's position: uses serving distinct purposes may warrant separate fair use analyses, but uses serving the same transformative purpose are evaluated together.

Here, the unrebutted evidence confirms that all of the data OpenAI downloaded was for the transformative purpose of training its LLMs. OSUF ¶ 104. Plaintiffs point to no evidence to the contrary. They concede that the purported distributions they identify were for training. *See* Sections II.E, III.A, *supra*. They speculate that the way OpenAI stored its training data would have allowed for uses other than training, but they identify no such uses. *See* Class Mot. at 11, 22. They cite OpenAI's internal data-access policy governing access to internal company information generally, including confidential technical materials, user data, and training data. *Id.* at 22. But this policy took effect on June 2, 2023—long after OpenAI had stopped using LibGen data—and says nothing about why OpenAI downloaded books years earlier or the use of the training data. Nor does it matter whether every downloaded book was ultimately used to train an In-Scope Model. *Contra* Class Mot. at 22. The purpose of acquiring training data does not change merely because some data is later excluded from training. *Kadrey*, 788 F.Supp.3d at 1048

(rejecting similar argument regarding use of LibGen data).[7] Unlike in *Bartz*, there is no basis to conclude that OpenAI used the Asserted Works for any purpose other than model training.

Nor does *Bartz* establish that the mere act of downloading from a "pirate library" is categorically excluded from fair use.[8] Although the *Bartz* court distinguished *Sega* and *Connectix* on the ground that the initial copies were "purchased commercially," it identified no statutory basis for treating the provenance of a copy as itself establishing separate uses for purposes of Section 107. *See Bartz*, 787 F.Supp.3d at 1028. Indeed, in other fair-use cases where the defendant acquired and used an unauthorized copy of the plaintiffs' work, courts have not analyzed the initial acquisition separately from the overall purpose for which the defendant used the work. *See*, *e.g.*, *NXIVM Corp. v. Ross Inst.*, 364 F.3d 471, 476 (2d Cir. 2004); *Perfect 10*, 508 F.3d at 1164 n.8; *Kadrey*, 788 F.Supp.3d at 1048. The source of a copy does not answer the fair use question; Section 107 requires examining what the defendant actually did with the work and why.

Class Plaintiffs' atomized fair use analysis is not driven by anything in the law, but instead by their desired outcome. Notably, the News Plaintiffs, represented by the same firm as Class Plaintiffs, have taken an entirely different approach to suit their litigation strategy in the News cases. While Class Plaintiffs insist that the court should analyze each prima facie act of

[7] The fact that Google scanned texts that it eventually ended up excluding from the Google Books tool did not make that scanning any less fair—because the scanning was part and parcel of Google's ultimate transformative purpose. *See* Decl. of Dan Clancy ¶ 12, *Authors Guild v. Google*, No. 05-cv-08136, ECF 1035 (Google scanned dictionaries but excluded them from search and snippet functionality).

[8] Class Plaintiffs rely heavily on a statement from *Bartz* regarding piracy. Class Mot. at 20–21 (quoting *Bartz*, 787 F.Supp.3d at 1025). That sweeping statement was explicitly *dicta*. *See Bartz*, 787 F.Supp.3d at 1025–26 (acknowledging that "this order need not decide this case on that rule").

infringement separately, News Plaintiffs insist that the Court need only conduct a single fair use analysis, even though their claims implicate not just the same set of multiple internal copies, but two distinct purposes altogether—namely, the use of works to train models and the use of works in the retrieval-augmented generation ("RAG") process. News Mot. at 50–74. The Court should reject both outcome-driven approaches in favor of the one dictated by the law, in which separate fair-use analyses are applied to distinct uses based on the purpose of those uses.[9]

B. OpenAI's Alleged Use Was Fair Use

OpenAI's alleged use of the Asserted Works was highly transformative, did not display any meaningful portion of the books to ChatGPT users, and caused no cognizable harm.

1. Factor One: OpenAI's Use of the Asserted Works Was Highly Transformative

The first factor considers "the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes." 17 U.S.C. § 107(1). "The central question it asks is whether the new work merely supersedes the objects of the original creation (supplanting the original), or instead adds something new, with a further purpose or different character." *Warhol*, 598 U.S. at 509 (cleaned up). The "larger the difference" between the purpose or character of the new use and the original work, "the more likely the first factor weighs in favor of fair use." *Id.* at 529. "A use that has a further purpose or different character is said to be transformative." *Id.* "The commercial nature of a use is relevant, but not dispositive." *Id.* at 510.

[9] Even if the Court were to accept Plaintiffs' flawed attempt to atomize the fair use analysis, the downloads and controlled transfers they complain about each were performed for the purpose of training and were fair use for the reasons described herein. OpenAI's purpose was fundamentally different from the purpose for which Plaintiffs authored the Asserted Works and Plaintiffs suffered no cognizable harm.

There is no genuine dispute that OpenAI's training use was "highly," "quintessentially," and "spectacularly" transformative. *Kadrey*, 788 F.Supp.3d at 1044; *Bartz*, 787 F.Supp.3d at 1021–22, 1026. Plaintiffs wrote their books to be read and enjoyed by a human audience. OSUF ¶ 95. OpenAI's alleged use of books was for a fundamentally different purpose—to train LLMs, which are innovative, general-purpose tools capable of interpreting and outputting natural language and performing a wide variety of functions. OpenAI's research papers, the testimony of OpenAI witnesses and experts, and the operation of OpenAI's LLMs all confirm that the purpose of pretraining is to teach the model generalizable patterns of language, not to reproduce protected expression. *See, e.g.*, OSUF ¶¶ 46–48. As the *Bartz* court observed, this technology is "among the most transformative many of us will see in our lifetimes." 787 F.Supp.3d at 1033.

Unable to meaningfully dispute that fundamental difference in purpose, Plaintiffs instead argue that OpenAI designed its product to have the capability to generate competing outputs and that its alleged copying was not "reasonably necessary" to its transformative purpose. Class Mot. at 26–33. These arguments fail as a matter of law.

a. Plaintiffs' Argument that OpenAI Designed its Products to be Capable of Generating Competing Outputs is Incorrect and Irrelevant

Plaintiffs contend that OpenAI designed its LLMs to write books that would compete with their books. Class Mot. at 26–27. As a preliminary matter, this is indisputably not the purpose of the training process. The In-Scope Models were not designed to (and are not capable of) generating full books, and none of the evidence Plaintiffs cite is to the contrary. Moreover, Plaintiffs' argument fails as a matter of law: the possibility that a technology may enable the creation of competing works does not make the copying undertaken to develop that technology nontransformative.

Sega and *Connectix* squarely reject Plaintiffs' theory. In both cases, defendants copied copyrighted video game code without authorization to develop new, non-infringing products that competed with the copyright owners' products.[10] *Sega*, 977 F.2d at 1522; *Connectix*, 203 F.3d at 606–07. Both courts found that this use was transformative, and the *Sega* court emphasized that copying that "facilitat[ed] the entry of a new competitor" affirmatively favored fair use. *Sega*, 977 F.2d at 1523; *see also Connectix*, 203 F.3d at 606–07 (competing work that does not contain plaintiff's protected expression was a "legitimate competitor in the market").

Those decisions establish a critical distinction: competition from a new, non-infringing product is not the kind of "substitution" copyright guards against. That principle extends far beyond the factual context of those cases (*i.e.*, reverse engineering of software) and is fully consistent with Second Circuit precedent. *See, e.g.*, *Arica Inst., Inc. v. Palmer*, 970 F.2d 1067, 1078 (2d Cir. 1992) ("Defendant, after all, is perfectly entitled to create a competing work."); *Google Books*, 804 F.3d at 224. More broadly, it also follows from the Supreme Court's admonition that fair use requires a searching inquiry not only of the impact of the use on a plaintiff, but also the "source of [that impact]," *Oracle*, 593 U.S. at 35, and, in particular, whether it constitutes the kind of "market substitution" that lies at the heart of fair-use analysis, *Campbell*, 510 U.S. at 591–93; *see also Google Books*, 804 F.3d at 224 ("loss of sale[s]" resulting from provision of "historical fact" does not constitute substitution or preclude a finding of transformativeness). Thus, even if new, non-infringing books created with the assistance of an

[10] As discussed in Section III.B.4.a, *infra*, Plaintiffs complain about competition with hypothetical, non-infringing books written by users of ChatGPT—not competition with OpenAI's products. This is even more attenuated than the competition at issue in *Sega* and *Connectix*.

LLM compete with Plaintiffs' books, that competition does not establish that OpenAI's training use was nontransformative.

Plaintiffs' reliance on *Warhol* ignores that distinction. Class Mot. at 27, 28. Unlike *Sega* and *Connectix*, *Warhol* was a case in which the claimed "substitution" occurred as a result of the defendant's re-publication of the plaintiff's protected expression—in particular, its use of that expression in a Warhol screen print that it licensed to a magazine. *Warhol*, 598 U.S. at 515–16. It was that kind of "substitution," via the use of protected expression, that *Warhol* dubbed copyright's "bête noire." *Id.* at 528. The case, in other words, did not implicate the distinction between "substitution" as a result of the use of expression and "competition" arising from the creation of new, non-infringing works. Nothing in *Warhol* addresses, let alone purports to abrogate, the principle embraced in *Sega*, *Connectix*, and *Google Books* that is dispositive here: using copyrighted content as part of the internal process of creating a new technology can be transformative even if the technology ultimately creates competition for the plaintiff's content. Moreover, the Supreme Court in *Warhol* approvingly cited *Google Books* three times, 598 U.S. at 531, 545, 548—just two years after approvingly citing both *Sega* and *Connectix* twice in *Oracle*, 593 U.S. at 22, 39.

The type of substitution at issue in *Warhol*—reproduction in the accused work of the expressive content of the original work—is ***not*** at issue in this case. Plaintiffs make no showing that pretraining results in regurgitation of substantial amounts of protected expression from their works to the public. They fail to address that OpenAI implemented numerous safeguards to ensure that its models could serve their transformative purpose without providing users any meaningful amount of expressive content from Plaintiffs' books. *See* OSUF ¶¶ 58–81.

Plaintiffs also fail to address testimony from their own experts (and OpenAI experts) showing that OpenAI's LLMs produce text that matches text from the Asserted Works extremely rarely and, when it occurs, the amount displayed is a tiny fraction of the book. OSUF ¶¶ 82–84, 91. Even after inputting over 5.3 million prompts designed to force a model to output verbatim text matching text from the Asserted Works, the longest span of contiguous text that Plaintiffs' expert could generate was a 1,899-word excerpt from *A Game of Thrones* that represents only 0.62% of the book. OSUF ¶¶ 85–89. *See supra*. OpenAI's models, in other words, are not designed to reveal any protected expression at all, and only do so extraordinarily rarely. Plaintiffs are therefore wrong to suggest that OpenAI's use raises more "risk of serving as a competing substitute" than the service at issue in *Google Books*. Class Mot. at 27. In fact, OpenAI's models pose far ***less*** of a risk of substitution than the Google Books service, which was specifically designed to show segments of the books' expressive material—by some measures, up to 16% of a book. *Google Books*, 804 F.3d at 222–23. That is far more than the longest span of text that Plaintiffs' experts could forcibly extract from OpenAI's LLMs (0.62%). *See* OSUF ¶¶ 85–89; *see also Google Books*, 804 F.3d at 222–23 (evaluating whether service is a "meaningful substitute" by evaluating quantum and manner in which it displayed protected expression). OpenAI's safeguards, the rarity and limited extent of observed regurgitation, and the results of Plaintiffs' extensive extraction attempts show that OpenAI's LLMs do not provide a meaningful substitute for Plaintiffs' books. If the deliberate display of copyrighted text in *Google Books* did not create a meaningful substitute, Plaintiffs cannot establish one here.

Plaintiffs attempt to distinguish *Kadrey* and *Bartz* by incorrectly arguing that those cases "lack[ed] the evidence of substitutive purpose here." Class Mot. at 28. Once again, OpenAI's purpose has nothing to do with reproducing protected expression and is therefore not the

"substitutive purpose" the Court considered in *Warhol*. Regardless, both *Bartz* and *Kadrey* rejected arguments similar to the argument Plaintiffs make here. *See Bartz*, 787 F.Supp.3d at 1017 (factor one favored fair use despite evidence that Anthropic trained on books to improve its model's ability to generate prose as accurate and compelling as the Plaintiffs' books); *Kadrey*, 788 F.Supp.3d at 1045 (factor one favored fair use even where "one possible use of Llama is to generate text with similarities to unprotectable aspects of the plaintiffs' books").

Thomson Reuters Enter. Centre GmbH v. ROSS Intel. Inc. does not dictate a contrary result for two reasons. ***First***, the Third Circuit explicitly distinguished this case and others finding that using copyrighted content to train Generative AI models is transformative, noting that "[u]nlike the AI models in *Bartz* and *In re: OpenAI*, ROSS's AI platform cannot generate original expression, and the evidence here supports the opposite conclusion about transformativeness." — F.4th —, 2026 WL 2932557, at *6 n.7 (3d Cir. Sept. 29, 2026) (finding that "ROSS trained its AI for the purpose of creating a commercial substitute for Westlaw").

Second, to the extent Class Plaintiffs read *ROSS* to hold that competition with a plaintiff by itself precludes transformativeness, that interpretation conflicts with controlling Second Circuit precedent, which recognizes that products may serve the same function or audience and still be transformative. *See, e.g.*, *Arica*, 970 F.2d at 1077–78 (competing self-help books; use was transformative); *Kane v. Comedy Partners*, 2003 WL 22383387, at *4 (S.D.N.Y. Oct. 16, 2003), *aff'd* 98 F.App'x 73 (2d Cir. 2004) (competing comedy TV shows; use was transformative); *Estate of Smith v. Cash Money Records, Inc.*, 253 F.Supp.3d 737, 749–50 (S.D.N.Y. 2017), *aff'd Estate of Smith v. Graham*, 799 F.App'x 36 (2d Cir. 2020) (competing songs; use was transformative). So too for *ROSS*'s suggestion that the fact that the plaintiff and the defendant are competitors necessarily establishes cognizable market harm under factor four. *Compare ROSS*,

2026 WL 2932557, at *8–10, *with Arica*, 970 F.2d at 1078 (fact that plaintiffs' books appealed to the same audience not a "relevant market effect" because "[d]efendant, after all, is perfectly entitled to create a competing work").

b. Plaintiffs Misinterpret the "Reasonable Necessity" Inquiry

Plaintiffs also argue that, under *Warhol*, OpenAI's use of copyrighted works cannot be fair use because the copying was not "reasonably necessary" to achieve the transformative purpose of training LLMs. *See* Class Mot. at 29–32. Because OpenAI could have used other books (or no books) to train its LLMs, Plaintiffs argue, OpenAI's copying cannot be justified and is not transformative. *See id.* at 30–31. This argument again misreads *Warhol*.

In *Warhol*, the Court addressed the concept of "reasonable necessity" in addressing the relevance of the defendant's "justification for the use." 598 U.S. at 531. It explained that, in general, "a use that has a distinct purpose is justified because it furthers the goal of copyright, namely, to promote the progress of science and the arts, without diminishing the incentive to create." *Id.* In that sense, any requirement to establish a "justification for the use" is satisfied when the defendant's purpose was sufficiently distinct from the original. Here, as explained, the purpose is spectacularly transformative. *See* OpenAI-Class Mot. at 20–23.

Warhol did not hold or even suggest that every fair user must demonstrate that using the plaintiff's specific copyrighted work was strictly "necessary" for their transformative purpose. Instead, the Court explained that "where an original work and copying use share ***the same or highly similar purposes***, or where wide dissemination of a secondary work would ***otherwise run the risk of substitution*** for the original or licensed derivatives of it," a secondary user must provide an "independent justification" for the use. 598 U.S. at 532. As discussed, the facts of that case were that the use in question ***did*** substitute for the protected expression in the original. In light of those facts, some particularly compelling justification for the use was required. *Id.* at

537–38. But *Warhol* did not purport to announce a new rule that no use can be fair absent the satisfaction of a strict "necessity" criterion.

In fact, the Second Circuit has already disposed of Plaintiffs' argument. The post-*Warhol* decision in *Romanova v. Amilus Inc.*, 138 F.4th 104 (2d Cir. 2025), explained that the universe of cognizable justifications is not limited to "necessity" or "commentary on the original." *Id.* at 115. To the contrary, *Romanova* catalogued a broad and diverse taxonomy of justifications that courts have recognized. *See id.* at 115–18.

Regardless, Plaintiffs' understanding of the "reasonable necessity" rubric is wrong even on its own terms. In *Google v. Oracle*, the Supreme Court held that Google's reimplementation of the Java SE API declarations was fair use, even though Google concededly ***could*** have written its own API from scratch or chosen a different API to copy. 593 U.S. at 27–29. The Court assessed whether the copying served a transformative purpose—enabling creative new programs on a new platform—and concluded that it did, without requiring Google to exhaust all non-infringing alternatives. *Id.* at 35. Similarly, in *Google Books*, Google could have created an adequate book search product without the specific books of the plaintiffs there. *See* 804 F.3d at 208. But the court did not require Google to prove that the use of plaintiffs' ***specific works*** was necessary. The question is whether the copying was justified by the transformative purpose it served, not whether that purpose could conceivably have been accomplished without Plaintiffs' particular books.

c. Commerciality Does Not Tip Factor One Against Fair Use

For the reasons stated in OpenAI's motion for summary judgment, OpenAI-Class Mot. at 25–26, commerciality does not tilt this factor against fair use, particularly in light of the transformative nature of OpenAI's use. *See Oracle*, 593 U.S. at 32; *Campbell*, 510 U.S. at 583.

d. Plaintiffs' Assertions Regarding "Piracy" Do Not Change the Fair Use Analysis

Plaintiffs devote much of their brief to arguments about "piracy." These arguments miss the mark. "To say that [OpenAI's] downloading was 'piracy' and thus cannot be fair use begs the question because the whole point of fair use analysis is to determine whether a given act of copying was unlawful." *Kadrey*, 788 F.Supp.3d at 1046.

To the extent Plaintiffs argue it can never be fair use to use content acquired from an "unauthorized" source, Class Mot. at 19, that argument clearly fails. Both the Second and Ninth Circuits have so held. *See*, *e.g.*, *NXIVM*, 364 F.3d at 475–76 (fair use factors favor defendant notwithstanding evidence that defendant's "access to the copyrighted materials was unauthorized"); *Perfect 10*, 508 F.3d at 1164 n.8 (finding fair use where Google used images "derived from infringing websites"); *see also Kadrey*, 788 F.Supp.3d at 1046, 1050 (granting summary judgment of fair use where defendant trained its LLMs on books downloaded from LibGen); *Bartz*, 787 F.Supp.3d at 1033 (same for copies Anthropic made to train its LLMs); *compare* Brief for Plaintiffs-Appellees, *Sony Comput. Ent. Inc. v. Connectix Corp.*, 1999 WL 33623858, at *9, 30 (9th Cir. June 24, 1999) (arguing that Connectix's use of "unauthorized copies" weighed against fair use), *with Connectix*, 203 F.3d at 602–08 (finding fair use).

Plaintiffs may argue that use of books from alleged "pirate websites" or "shadow libraries" is "bad faith" conduct that should be considered as part of factor one. But the Supreme Court has repeatedly counseled that "bad faith" should be irrelevant. *See Campbell*, 510 U.S. at 585 n.18 (expressing skepticism about whether bad faith has any role in a fair use analysis); *Oracle*, 593 U.S. at 32 (reiterating the Court's skepticism and quoting Judge Leval's admonition that "[c]opyright is not a privilege reserved for the well-behaved"). And for good reason: as Judges Jacobs and Leval have separately explained, fair use "defines the outer boundary of

copyright protection," such that the "perimeter should be drawn by reference to the central objectives of copyright" rather than "the secondary user's good or bad faith." *NXIVM*, 364 F.3d at 483–86 (Jacobs, J., concurring); *see also* Pierre N. Leval, *Campbell as Fair Use Blueprint?*, 90 Wash. L. Rev. 597, 612–13 (2015) ("A copier's bad faith has no logical bearing on the scope of the original author's copyright."). In any case, even the opinions considering "bad faith" as a relevant criterion afforded it minimal weight. *See, e.g.*, *NXIVM*, 364 F.3d at 479 & n.2. The Second Circuit has held that it "cannot be made central to fair use analysis." *Id.*

Moreover, the undisputed factual record refutes any inference of "bad faith." Plaintiffs concede (1) the downloads they complain about occurred years ago, before OpenAI had any commercial products and long before ChatGPT was released, CSUF ¶¶ 105–106, 145, 193; (2) the downloads were for the purpose of training LLMs, CSUF ¶¶ 188, 189, 195–200; (3) the only In-Scope Models that were trained on books downloaded from LibGen are GPT-3 and GPT-3.5, CSUF ¶ 519; and (4) OpenAI "stopped using LibGen to train its publicly released GPT models in late 2021," CSUF ¶ 265. Plaintiffs present no evidence that OpenAI provided support—financial or otherwise—for pirate websites or shadow libraries. Their conclusory assertion that OpenAI "normaliz[ed]" or "encouraged" use of these sources is based on the *ipse dixit* of their expert and a misleading citation to a statement from a different shadow library (Anna's Archive) from which Plaintiffs do not even allege OpenAI downloaded books at all. Class Mot. at 24.

2. Factor Two: OpenAI's Alleged Copying Was Not to Capture or Republish Protected Expression

Factor two considers "the nature of the copyrighted work[.]" 17 U.S.C. § 107(2). It "has rarely played a significant role in the determination of a fair use dispute." *Google Books*, 804 F.3d at 220; *accord Lennon v. Premise Media Corp.*, 556 F.Supp.2d 310, 325 (S.D.N.Y. 2008).

OpenAI did not use the Asserted Works to republish the creative expression in them. Instead, OpenAI used billions of textual works to access unprotectable learnings from and elements in those works. OSUF ¶¶ 60–61; *see also* OpenAI-Class Mot. at 26–27.

3. Factor Three: OpenAI Used a Reasonable Amount of the Asserted Works in Light of Its Transformative Purpose

Factor three considers "the amount and substantiality of the portion used in relation to the copyrighted work as a whole." 17 U.S.C. § 107(3). Plaintiffs argue this factor weighs "strongly against fair use" because OpenAI allegedly "copied the entirety of the works." Class Mot. at 23, 33. But copying an entire work does not preclude fair use. To quote *Bill Graham Archives v. Dorling Kindersley Ltd.*: "copying the entirety of a work is sometimes necessary to make a fair use of the [work]." 448 F.3d 605, 613 (2d Cir. 2006) (third factor "does not weigh against fair use" even though defendant "reproduced [the works] in their entirety"). Here, it is undisputed that pretraining an LLM requires an enormous volume of text so the model can distill broad patterns that emerge across hundreds of billions of words. Accordingly, OpenAI's alleged use of entire books is reasonable in light of that transformative purpose, and the third factor weighs in favor of fair use. *See id.* at 613.

Importantly, the Second Circuit held in *Google Books* that "[w]hat matters" is not the amount of the work used "in making a copy," but "rather the amount and substantiality of ***what is thereby made accessible*** to a public for which it may serve as a competing substitute." 804 F.3d at 222. OpenAI's LLMs are not designed to reproduce ***any*** of Plaintiffs' protected expression to the public. And in the exceedingly rare instances of regurgitation, the amount of expression displayed is far too short to constitute a "significantly competing substitute for the copyrighted work." *Id.* at 222; OSUF ¶ 82, 90; *see also Bartz*, 787 F.Supp.3d at 1029–30 (third

factor favors fair use, including because outputs do not infringe); *Kadrey*, 788 F.Supp.3d at 1050 (similar).

Courts have likewise afforded the third factor little weight where the use serves a transformative technological purpose. *See Connectix*, 203 F.3d at 606 (according factor three "very little weight"); *Sega*, 977 F.2d at 1526–27 (same); *see also Google Books*, 804 F.3d at 221–22. Thus, even if the court were to find that the use of entire books weighed against fair use, it would not be significant here given OpenAI's transformative purpose.

4. Factor Four: OpenAI's Alleged Use Did Not Cause Cognizable Harm

The fourth factor considers the effect the copying has on the "potential market for or value of the copyrighted work." 17 U.S.C. § 107(4). OpenAI has shown that Plaintiffs have not suffered any cognizable market harm under factor four. OpenAI-Class Mot. at 28–39. Plaintiffs fail to make any contrary showing. They fail to identify any tangible effect on the market for or value of the Asserted Works. They do not claim lost sales, declining prices, or diminished market share. They present no analyses, studies, or surveys showing that any such harm is likely to occur in the future. Instead, they assert they "need not offer empirical data" to show harm, and they speculate that some authors—though not necessarily these Plaintiffs—might be harmed in the future because people will choose to buy non-infringing "AI-generated" books rather than human-authored books. Class Mot. at 33–36. These theories of harm are both legally non-cognizable and unsupported by the record.

a. "Market Dilution" is Not a Cognizable Theory of Harm

Plaintiffs' primary fourth-factor argument rests on a theory of "market dilution." Relying on *dicta* from *Kadrey*, they claim that "AI-generated" books are flooding the market and creating more competition for the Asserted Works or books Plaintiffs might write in the future. Class Mot. at 34–36. This speculative theory of harm is not just factually unsupported, but also legally

wrong, and would, if adopted, fundamentally reshape the fair use doctrine in ways the Supreme Court has rejected.

The fourth-factor inquiry focuses on the kind of substitution that occurs when a defendant republishes protected expression from the plaintiff's work, thereby raising "the likelihood that potential purchasers may opt to acquire the [defendant's] copy in preference to the original." *Google Books*, 804 F.3d at 223; *accord Campbell*, 510 U.S. at 590. This rule has two relevant implications: (1) alleged market harm resulting from competition with non-infringing works is not cognizable as a matter of law; and (2) the defendant's use must ***itself*** displace demand for the protected elements of the original—not merely enable the creation of a new technology that might be used by third-parties to create non-infringing works that compete in the same market as the original.

First, to the extent Plaintiffs claim harm from competition with new books that "aren't themselves infringing," that market dilution theory is untethered from copyright law and at odds with Second Circuit precedent. Class Mot. at 34. The Second Circuit has made clear that the only "effect on the market" that is relevant to fair use is the effect caused by the use of the plaintiff's "expression." *Wright v. Warner Books, Inc.*, 953 F.2d 731, 739 (2d Cir. 1991). The mere fact that a non-infringing work addresses the same general subject matter, or generally appeals to the same audience, is not relevant. *Id.* (discounting market effects resulting from "[d]isclosure of factual content"); *see also Google Books*, 804 F.3d at 224 (similar). As the Second Circuit ruled in *Arica Institute, Inc. v. Palmer*, "the relevant market effect is that which stems from defendant's use of plaintiff's 'expression,' not that which stems from defendant's work as a whole." 970 F.2d 1067, 1078 (2d Cir. 1992); *see also* 4 Patry on Copyright § 10:150 ("[E]ven though the taking of unprotectable material such as important ideas, laboriously researched facts,

or the copying of the work's overall style may cause lower sales, such losses should not be considered. The harm must be caused by the use of expression."); *Monster Commc'ns, Inc. v. Turner Broad. Sys., Inc.*, 935 F.Supp. 490, 495–96 (S.D.N.Y. 1996) (holding that the "only effect of which plaintiff properly may complain is that caused by the alleged infringement" and declining to consider market effects of non-infringing aspects of competing film).

Accordingly, general competition between a defendant and plaintiff, standing alone, has never sufficed to establish market harm. The Second Circuit stated this explicitly in *Arica*. 970 F.2d at 1078 (dismissing the fact that defendant's book "might well interest" plaintiff's audience as irrelevant; finding no market harm). Other circuits have held the same. *See, e.g.*, *Connectix*, 203 F.3d at 607 (loss of sales from "a legitimate competitor in the market" does not weigh against the defendant); *Sony Comput. Ent. Am., Inc. v. Bleem*, *LLC*, 214 F.3d 1022, 1029 (9th Cir. 2000), *amended on denial of reh'g* (July 10, 2000) (holding that even the "los[s] [of] a significant share of [the plaintiff's] present market" is not cognizable if it resulted "not from the display of plaintiff's [expressive material] but from commercial competition with a work that does not in any way make use of plaintiff's copyrighted material"); *Triangle Publ'ns, Inc. v. Knight-Ridder Newspapers, Inc.*, 626 F.2d 1171, 1177 (5th Cir. 1980) (holding that loss of market share from "commercial competition with a work that does not in any way make use of plaintiff's copyrighted material" does not weigh against fair use).

Plaintiffs' market-dilution argument relies principally on *dicta* in *Kadrey*, 788 F.Supp.3d at 1052–54, speculating about "indirect substitution" that might result from the "flooding" of stores with non-infringing new books. But as discussed above, the Second Circuit and others have long held that the only "relevant market effect" that concerns the fourth factor is "that which stems from defendant's use of plaintiff's 'expression.'" *Arica*, 970 F.2d at 1078; *see also*

Wright, 953 F.2d at 739; *Google Books*, 804 F.3d at 224–25. Harm arising from the distribution of new books that do not contain any of Plaintiffs' expression is not harm "caused by the use of expression." 4 Patry on Copyright § 10:150. As *Bartz* held, "[t]his is not the kind of competitive or creative displacement that concerns the Copyright Act." 787 F.Supp.3d at 1031–32.

Indeed, for these reasons, this *dicta* in *Kadrey* has been sharply criticized by leading commentators. 4 Patry on Copyright § 10.155.40 (criticizing the "market dilution" theory as contrary to "the 300-year old requirement . . . that harm to the market must be harm to the individual work"); *see also* MDL ECF 1682 ("Statement of Interest of the USA") at 16 (explaining that *Kadrey*'s *dicta* "overlooks arguably the most basic element of copyright infringement" by recognizing harm from LLM outputs that are not "substantially similar" to a plaintiff's work).

Second, the non-cognizable harm that Plaintiffs complain about (books created by third-parties) is far removed from the use they allege to be infringing (OpenAI's use of their books to pretrain LLMs). OpenAI does not publish books that compete with Plaintiffs. Rather, OpenAI built a general-purpose LLM that can be used by others for a wide variety of tasks, one of which is writing. In fact, Plaintiffs themselves use OpenAI's LLMs to help them with their writing and acknowledge that many other authors do, too. OSUF ¶¶ 133, 135; Nelson Decl. Ex. 90. No controlling authority has ever found cognizable harm in these circumstances. Indeed, the fourth-factor inquiry focuses on the direct effect of the defendant's use on the "market for plaintiffs' 'expression,'" not on speculation regarding the indirect effects of copying on the general marketplace for works of the type that Plaintiffs have created or might create in the future.[11] *NXIVM*, 364 F.3d at 482.

[11] Indeed, as noted below Plaintiffs claim "market dilution" harm even as to Asserted Works that they concede OpenAI did not use in training. *See infra* n.12. This further demonstrates that the

Google Books further illustrates the point. In analyzing whether Google's display of snippets of text from copyrighted books was a fair use, the Second Circuit considered whether the snippet function itself "provide[d] searchers with any meaningful experience of the expressive content of the book." 804 F.3d at 226–27. Notably, the Court did not entertain farther-reaching theories of harm, like whether Google Books would enable aspiring authors to research and write new books that would compete in the marketplace. *See id*. Likewise, in *Google v. Oracle*, the Supreme Court found that Google's use of Oracle's APIs was fair use because it enabled Android users to create new applications. 593 U.S. at 39–40. It did not ask whether those new applications would compete with Oracle's existing or future business. *Id.*

To extend the fourth-factor harm analysis to encompass speculative market impacts that might flow from non-infringing books created by some tiny fraction of OpenAI's users would be tantamount to saying that it can never be a fair use to copy books to create a general-purpose technology capable of writing, irrespective of its other capabilities. The exclusive rights granted to copyright owners cannot stretch so far.

b. Plaintiffs Have Not Suffered *Any* Harm, Cognizable or Not

In any case, OpenAI has shown that Plaintiffs have not suffered harm from "market dilution," and, after years of discovery, Plaintiffs fail to present any contrary evidence. OpenAI's economic expert, Dr. Gregory Leonard, conducted five distinct analyses of sales and pricing data across the trade book industry and for the Asserted Works. OSUF ¶¶ 125–129. All five analyses reached similar conclusions—the release of OpenAI's LLMs via OpenAI's API and ChatGPT was not associated with a systematic change in the sales or prices of the books being analyzed. *Id.* Plaintiffs do not dispute this with any empirical analysis of their own, and they present no

non-cognizable harm that Plaintiffs' claim is disconnected from the infringement they allege.

evidence of actual harm. In fact, their factor-four expert claimed he did not have the "data necessary" to do so. OSUF ¶ 131.

Plaintiffs brush off this deficiency by arguing that they need not offer empirical data because the Court can simply resort to "logical inferences." Class Mot. at 33–34 (quoting *Hachette*, 115 F.4th at 193). But Dr. Leonard's analysis directly refutes the inference Plaintiffs seek to draw. OSUF ¶¶ 125–129 (discussing Dr. Leonard's finding that OpenAI's LLMs had no impact on sales and pricing across the trade book industry and for the Asserted Works). Moreover, "[n]o . . . inference of market harm . . . is applicable to a case," as here, that "involv[es] something beyond mere duplication for commercial purposes." *Campbell*, 510 U.S. at 591. Because "the second use is transformative, market substitution is at least less certain, and market harm may not be so readily inferred." *Id.*

"While it made sense to infer market harm in *Hachette*, it doesn't make sense to do so here." *Kadrey*, 788 F.Supp.3d at 1057. In *Hachette*, the defendant "create[d] digital copies of the [plaintiff's] Works and distribute[d] those copies to its users in full, for free." *Hachette*, 115 F.4th at 181. Here, OpenAI does not make copies of the Asserted Works available to the public. *See Kadrey*, 788 F.Supp.3d at 1057 (declining to infer harm because defendant's "use does not let users access any significant portions of the plaintiffs' books"). Plaintiffs present no reason why the court should simply assume that, for example, the market for or value of John Grisham's *The Pelican Brief*—a bestselling book that has been made into a popular movie—has been or will be impacted by unidentified, non-infringing, AI-generated books that may or may not exist yet. *See Kadrey*, 788 F.Supp.3d at 1052 (explaining that it "seems unlikely" that "AI-generated books would meaningfully siphon sales away from well-known authors."). Indeed, Mr. Grisham himself testified that he is not aware of "any instance where [he] lost a sale of a book in

competition with a ChatGPT-generated book" and "can't identify a single reader who chose not to purchase one of [his] books because they had a ChatGPT generated book instead." OSUF ¶ 116.

Moreover, it strains credulity for Plaintiffs—best-selling, award-winning authors who have sold hundreds of millions of books—to speculate that they have been or will be harmed by competition from unidentified, non-infringing, AI-generated books, while at the same time asserting that AI-generated works could never match the quality of their own books. *See* CSUF ¶ 857 ("AI-containing books garner substantially less usage per title and receive lower star ratings."); RCSUF ¶ 976 ("Well, I said earlier that [ChatGPT] wouldn't replace writers. What I mean is it can't replace – I don't think a machine-generated character-driven story will carry the heart and soul of something that comes from a human."); RCSUF ¶ 977 ("the current versions [of OpenAI's LLMs] are risibly awful at generating prose that I might have written."). It also contradicts the evidence, which shows that the In-Scope Models "cannot write long-form narratives." Nelson Decl. Ex. 764 at 3; *see also* Nelson Decl. Ex. 762.

Plaintiffs' evidence of AI-generated books on [redacted] does not establish harm to the Asserted Works. Class Mot. at 36; CSUF ¶¶ 849–854. [redacted]

Plaintiffs' cited evidence not only fails to trace any market harm to the Asserted Works, but even as to books in general it shows that "the AI influx does not appear to displace the volume or nature of human book authorship." *See* Nelson Ex. 762 at 3. The only evidence Plaintiffs cited that purported to suggest otherwise was the Chakrabarty paper, which Plaintiffs later withdrew after OpenAI moved to strike it based on undisclosed evidence that it was funded by Plaintiffs' litigation counsel. *See* CSUF ¶¶ 858–872; MDL ECF 2087 at 2; MDL ECF 2090.[12]

c. Plaintiffs Have Not Suffered Harm to Any Cognizable Licensing Market

Plaintiffs next argue they have been deprived of licensing revenues from an "existing, growing market to license works for LLM training." Class Mot. at 36. This theory of harm fails for two reasons. ***First***, it is not legally cognizable. *See Oracle*, 593 U.S. at 38. "[I]t is a given in every fair use case that plaintiff suffers a loss of a ***potential*** market if that potential is defined as the theoretical market for licensing the very use at bar." *Id.* (quoting Nimmer on Copyright § 13.05[A][4]). As the Second Circuit has held, alleged harm from a lost licensing fee in a "transformative market" is not cognizable as a matter of law. *See Bill Graham*, 448 F.3d at 615; *HathiTrust*, 755 F.3d at 100 ("[I]t is irrelevant that the Libraries might be willing to purchase licenses in order to engage in this transformative use (if the use were deemed unfair)."); *Google Books*, 804 F.3d at 226–27 (similar). For these reasons, both *Kadrey* and *Bartz* rejected similar arguments regarding a purported licensing market for training LLMs. *Kadrey*, 788 F.Supp.3d at

[12] Even if Plaintiffs had identified harm to an Asserted Work (they have not), there is no basis to assume that all 194 Asserted Works have been harmed. For 37 Asserted Works, Dr. Shan identifies no copy in the training data of any In-Scope Model. *See* Shan Decl. Ex. L, Tab "8-In-Suit Models Training Data." For the remainder, the alleged use varies across models—*e.g.*, some were allegedly used to train only GPT-3 or GPT-3.5, others only later models. *See id.* Plaintiffs do not even attempt to grapple with these differences and instead simply assume that every Asserted Work has been "displaced" or had its market "diluted" by AI-generated books.

1052 ("[T]his market is not one that the plaintiffs are legally entitled to monopolize."); *Bartz*, 787 F.Supp.3d at 1032 ("[S]uch a market for that use is not one the Copyright Act entitles Authors to exploit."). Plaintiffs fail to address any of these authorities. Instead, they cite cases in which the plaintiffs had an established revenue stream from licensing their works in a non-transformative market, where defendants interfered with that revenue stream by providing plaintiffs' works to end-users for free. *See*, *e.g.*, *Texaco*, 60 F.3d at 931; *Hachette*, 115 F.4th at 189–90; *UMG Recordings, Inc. v. MP3.Com, Inc.*, 92 F. Supp. 2d 349, 352 (S.D.N.Y. 2000).[13]

Second, even if harm to a transformative market were cognizable, no reasonable jury could find that there was or is a functioning licensing market in which Plaintiffs can license their Asserted Works for pretraining LLMs. Plaintiffs have never licensed their books for training LLMs, nor have they ever even been offered a licensing fee for that use. OSUF ¶¶ 121, 124; *White v. West Pub. Corp.*, 2014 WL 3385480, at *3 (S.D.N.Y. July 11, 2014) (finding no licensing market where a plaintiff had not received an offer to license his works, nor had he sought to license them). The Authors Guild's own contemporaneous documents make clear that a "licensing solution" did not exist as of July 5, 2023, after which the Guild explicitly set out to create a market to "help the litigation." RCSUF ¶¶ 971–977. Those efforts were unsuccessful, as the Authors Guild's corporate designees admitted in deposition. RCSUF ¶¶ 972, 973–974.[14]

[13] The Third Circuit's finding of licensing-market harm in *ROSS* does not bear on this legal rule. In that case, the court did not address whether the market at issue was a transformative market and it emphasized that Thomson Reuters was already using the headnotes for the purpose of training its own AI search products. *See* 2026 WL 2932557 at *6–7, 9.

[14] Even if they had been successful, it would not matter in this case. *See Bill Graham*, 448 F.3d at 614–15 ("[A] copyright holder cannot prevent others from entering fair use markets merely by developing or licensing a market for . . . transformative uses of its own creative work.") (cleaned up); *Texaco*, 60 F.3d at 931 (market analysis assessed based on the time of the relevant use).

Plaintiffs point to evidence showing that other content providers have entered into agreements with AI companies and that licensing aggregators have attempted to facilitate such agreements. Class Mot. at 37. But none of this evidence is informative of a market to license the Asserted Works, which are trade books. Of the successfully executed agreements, only one agreement involved access to trade books: [redacted]. But [redacted] RCSUF ¶ 978; *see also* MDL ECF 1707 ("Usdan Decl.") ¶¶ 22–25. Confronted with the failure of the [redacted] Usdan Decl. ¶ 25. Plaintiffs also point to OpenAI's partnership agreement with [redacted] but fail to acknowledge that the agreement provided access to [redacted]. *See* RCSUF ¶¶ 979–980. Further that agreement lasted [redacted], RCSUF ¶¶ 981–982, [redacted], RCSUF ¶ 983, [redacted]. RCSUF ¶ 984. [redacted]. RCSUF ¶ 985. In short, these agreements do not establish a functioning market for licensing books for pretraining an LLM.

In any event, the Second Circuit has twice found no licensing-market harm notwithstanding evidence that people (including the defendant) have entered into licensing agreements for the use in question. *See HathiTrust*, 755 F.3d at 100 ("[I]t is irrelevant that the Libraries might be willing to purchase licenses in order to engage in this transformative use (if the use were deemed unfair)."); *Google Books*, 804 F.3d at 226–27 (similar).

Plaintiffs also point to evidence of OpenAI's employees speculating on the value they would pay for access to books for training data. Class Mot. at 38. The Second Circuit has held, however, that "a publisher's willingness to pay . . . does not establish that the publisher may not, in the alternative, make fair use of [the Asserted Works]." *Bill Graham*, 448 F.3d at 615.

d. Alleged Lost Sales From the Uses at Issue are Not Cognizable and are Insufficient as a Matter of Law

Plaintiffs argue that they have been harmed because OpenAI did not purchase the books it downloaded for training or allegedly distributed to Microsoft and third-party contractors. Class Mot. at 23–24; 39. In Plaintiffs' view, they would be further harmed if these activities became widespread. *Id*. This argument fails as a matter of law.

The first aspect of Plaintiffs' argument—that "each digital book OpenAI copied for free deprives authors of the revenues due to them," Class Mot. at 23—rests on a harm too insubstantial to be counted. *See Google Books*, 804 F.3d at 224 (holding that "[t]here must be a meaningful or significant effect" on the market). As an initial matter, if the focus is on a single foregone purchase of each of the Asserted Works, OpenAI could have purchased all of the works on the used market, resulting in zero revenues to Plaintiffs. And a purchase of each Asserted Work new at full retail price would have yielded combined royalties to all named Plaintiffs amounting to at most a few hundred dollars. This is not a significant enough harm to "tilt the weighty fourth factor in favor of" Plaintiffs. *Id.* Viewed more generously, this argument is better understood as a restatement of Plaintiffs' general claim that copying from unauthorized sources can never constitute a fair use, a claim already refuted above.

The second aspect of Plaintiffs' argument—that widespread conduct like OpenAI's would mean that "anyone could lawfully obtain any book they wanted for free"—fares no better. The "widespread conduct" inquiry considers the impact of "unrestricted and widespread conduct

***of the sort engaged in by the defendant*[**.]" *Campbell*, 510 U.S. at 590; *Hachette*, 115 F.4th at 193; *see also Harper & Row Publishers, Inc. v. Nation Enters*, 471 U.S. 539, 568 (1985) (considering whether "the challenged use" would become widespread). The conduct at issue is not the download or distribution of pirated copies of books for human consumption. It is the download of books for the specific purpose of training LLMs, the secure distribution of works between parties to a contract for that purpose, and the restricted display of book excerpts to contractors to create training data for LLMs. *See* Section III.A, *supra*. When viewed in the correct light, there is no "logical inference[]" that Plaintiffs would suffer significant harm if the sort of conduct engaged in by OpenAI became widespread. *Contra* Class Mot. at 24.

C. Plaintiffs Cannot Assert an Unpleaded Distribution Claim

As noted above, Plaintiffs assert that OpenAI infringed their exclusive distribution right under 17 U.S.C. § 106(3) by "horse-trading" datasets containing Asserted Works and by providing Asserted Works to contractors. *See* Class Mot. at 38. Any such "distribution" was fair use as a matter of law (the exchanges were solely intended to further the training of LLMs, the same activities that constitute a fair use for the reasons already discussed), but, regardless, Plaintiffs should not be permitted to assert a distribution claim here because they failed to plead it in the Consolidated Class Action Complaint, *see* MDL ECF 183, and never sought leave to amend their pleadings to add such a claim. Plaintiffs cannot amend their pleadings through a summary-judgment motion. *Wright v. Ernst & Young LLP*, 152 F.3d 169, 178 (2d Cir. 1998) ("party may not amend pleading through statements in briefs"). Their distribution claim should therefore be denied in its entirety for this reason alone.

D. Plaintiffs Cannot Rely on Models Outside the Scope of this Case

In addition to asserting unpleaded claims, Plaintiffs assert claims related to out-of-scope models. These claims are subject to OpenAI's fair use defense, but they fail for the additional

reasons that they violate court orders. The Court has repeatedly held that Plaintiffs may not expand the scope of this case beyond the In-Scope Models. MDL ECF 707 at 1; *see* Section II.A n.2, *supra*. Yet several of the claims Plaintiffs assert in their Motion relate to out-of-scope research models:

Books3/The Pile. Plaintiffs assert OpenAI downloaded a copy of The Pile, but the partial copy of The Pile that Plaintiffs identify was used to train "very tiny" internal research models to evaluate The Pile's quality as training data—not to train any In-Scope Model. RCSUF ¶ 946.

Project Taxi. Plaintiffs claim that OpenAI infringed their copyrights by receiving portions of Bing Index data from Microsoft as part of "Project Taxi." CSUF ¶¶ 593–596. OpenAI used portions of Bing Index data to evaluate the data for potential use in future models, an evaluation conducted by training "internal-only research models," not any In-Scope Models. RCSUF ¶¶ 953–956.

Book-Summarization Project. Plaintiffs claim that OpenAI distributed some of the Asserted Works to contractors as part of its "book-summarization project," that project involved contractors viewing excerpts of books via a restricted web portal to conduct research on experimental training techniques, none of which was used to train In-Scope Models. CSUF ¶¶ 607–611; RCSUF ¶¶ 957–958.

IV. CONCLUSION

For the foregoing reasons, OpenAI respectfully requests that the Court deny Plaintiffs' motion. Instead, the Court should grant OpenAI's motion for summary judgment and find that OpenAI's use was fair use as a matter of law.

Dated: October 23, 2026

Respectfully submitted,

LATHAM & WATKINS LLP
Andrew M. Gass (*pro hac vice*)
andrew.gass@lw.com
505 Montgomery Street, Suite 2000
San Francisco, CA 94111
Telephone: 415.391.0600

Sarang V. Damle
sy.damle@lw.com
Luke A. Budiardjo
luke.budiardjo@lw.com
1271 Avenue of the Americas
New York, NY 10020
Telephone: 212.906.1200

Elana Nightingale Dawson (*pro hac vice*)
elana.nightingaledawson@lw.com
555 Eleventh Street, NW, Suite 1000
Washington, D.C. 20004
Telephone: 202.637.2200

MORRISON & FOERSTER LLP
Joseph C. Gratz (*pro hac vice*)
jgratz@mofo.com
Tiffany Cheung (*pro hac vice*)
tcheung@mofo.com
Caitlin Sinclaire Blythe (*pro hac vice*)
cblythe@mofo.com
425 Market Street
San Francisco, CA 94105
Telephone: 415.268.7000

Lily Li (*pro hac vice*)
YLi@mofo.com
755 Page Mill Road
Palo Alto, CA 94304
Telephone: 650.813.5600

KEKER, VAN NEST & PETERS LLP
Robert A. Van Nest (*pro hac vice*)
rvannest@keker.com
R. James Slaughter (*pro hac vice*)
rslaughter@keker.com
Paven Malhotra
pmalhotra@keker.com
Michelle S. Ybarra (*pro hac vice*)
mybarra@keker.com
Reid P. Mullen (*pro hac vice*)
rmullen@keker.com
Nicholas S. Goldberg (*pro hac vice*)
ngoldberg@keker.com
633 Battery St.
San Francisco, CA 94111
Telephone: 415.391.5400

Attorneys for OpenAI

CERTIFICATE OF COMPLIANCE

In accordance with Local Civil Rule 7.1(c), I hereby certify that the foregoing OPENAI'S OPPOSITION TO CLASS PLAINTIFFS' MOTION FOR SUMMARY JUDGMENT is 12,324 words, exclusive of the caption page, table of contents, table of authorities, and signature block. The basis of my knowledge is the word count feature of the word-processing system used to prepare this memorandum.

Dated: October 23, 2026 By: 