UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

In Re: OpenAI, Inc. Copyright Infringement Litigation This Document Relates To: Case No. 1:23-cv-08292 Case No. 1:23-cv-10211	Case No. 1:25-md-3143-SHS-OTW **REDACTED – PUBLIC VERSION**

Defendant Microsoft Corporation's Reply Memorandum Of Law In Support Of Motion For Summary Judgment In Books Plaintiffs' Consolidated Cases

TABLE OF CONTENTS

Page

TABLE OF AUTHORITIES ii
TABLE OF CITATION CONVENTIONS iv
INTRODUCTION 1
ARGUMENT 2
I. Training An LLM Is Fair Use As A Matter Of Law. 2
A. Plaintiffs fail to undermine the transformative purpose and character of LLM training (Factor one). 3
B. Books Plaintiffs fail to create a dispute on substitutive harm in traditional markets (Factor four) 12
II. Microsoft Is Entitled To Summary Judgment On Plaintiffs' Distribution Claims. 15
A. The distributions at issue were related to Defendants' collaboration in LLM training and development. 16
B. Defendants' transfers of data to facilitate LLM training are fair use. 19
1. The transfers were for the transformative purpose of LLM training. 19
2. Data transfers between partners in LLM development do not substitute in traditional markets for copyrighted works. 22
III. Microsoft Cannot Be Held Liable For OpenAI's Downloading From Shadow Libraries 24
CONCLUSION 25
CERTIFICATE OF WORD COUNT COMPLIANCE 29

TABLE OF AUTHORITIES

Page(s)

Cases

Am. Geophysical Union v. Texaco Inc.,
60 F.3d 913 (2d Cir. 1994) 22

Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith,
598 U.S. 508 (2023) 3, 8, 9

Argus Inc. v. Eastman Kodak Co.,
801 F.2d 38 (2d Cir. 1986) 15

Authors Guild, Inc. v. HathiTrust,
755 F.3d 87 (2d Cir. 2014) 3, 9, 10, 13, 23

Authors Guild v. Google, Inc.,
804 F.3d 202 (2d Cir. 2015) 7, 8, 16, 21

Campbell v. Acuff-Rose Music, Inc.,
510 U.S. 569 (1994) 13, 14

Castle Rock Ent., Inc. v. Carol Publ'g Grp., Inc.,
150 F.3d 132 (2d Cir. 1998) 13

Cox Commc'ns, Inc. v. Sony Music Ent.,
607 U.S. 583 (2026) 24

Google LLC v. Oracle Am., Inc.,
593 U.S. 1 (2021) 10, 11, 13

Graham v. Prince,
2023 WL 3383029 (S.D.N.Y. May 11, 2023) 10

Hachette Book Grp., Inc. v. Internet Archive,
115 F.4th 163 (2d Cir. 2024) 23

Kadrey v. Meta Platforms, Inc.,
788 F. Supp. 3d 1026 (N.D. Cal. 2025) 19

Maxtone-Graham v. Burtchaell,
803 F.2d 1253 (2d Cir. 1986) 5

N.Y. Times Co. v. Microsoft Corp.,
777 F. Supp. 3d 283 (S.D.N.Y. 2025) 8

Nat'l Basketball Ass'n v. Motorola, Inc.,
105 F.3d 841 (2d Cir. 1997) 8

Sega Enters. Ltd. v. Accolade, Inc.,
977 F.2d 1510 (9th Cir. 1992) 10

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

Thomson Reuters Enter. Centre GmbH v. Ross Intel. Inc.,
193 F.4th 271 (3d Cir. 2026) 12, 14

Whyte Monkee Prods., LLC v. Netflix, Inc.,
174 F.4th 761 (10th Cir. 2026) 9

Statutes

17 U.S.C. § 102(b) 7

TABLE OF CITATION CONVENTIONS

This brief uses the following abbreviations and citation conventions for briefs, Rule 56.1 statements, and record materials filed in connection with the parties' cross-motions for summary judgment in the News Plaintiffs' and Books Plaintiffs' cases.

Citation	Document	ECF No.
BP Msft Opp.	Class Plaintiffs' Memorandum of Law in Opposition to Microsoft's Motion for Summary Judgment	2260
BP OAI Opp.	Class Plaintiffs' Memorandum of Law in Opposition to OpenAI's Motion for Summary Judgment	2262
BP Resp-Msft SUMF ¶	Class Plaintiffs' Response to Microsoft's Rule 56.1 Statement of Undisputed Material Facts in Support of Microsoft's Motion for Summary Judgment	2213
BP SUMF ¶	Class Plaintiffs' Corrected Rule 56.1 Statement of Undisputed Material Facts in Support of Class Plaintiffs' Motion for Partial Summary Judgment	1988
Callison-Burch Opp., Ex. B ¶	Exhibit B to Declaration of Expert Chris Callison-Burch in Opposition to Books Plaintiffs' Motion for Summary Judgment	2121-2
Defs' Books Counter-SUMF ¶	Defendants' Response to Books Plaintiffs' Rule 56.1 Statement	2237
Gaggara ¶	Declaration of Padma Priya Gaggara in Support of Microsoft Corporation's Motions for Summary Judgment	1705
Hurst Ex.	Declaration of Annette L. Hurst in Support of Defendant Microsoft's Replies in Support of Microsoft's Motions for Summary Judgment	
Krishnan Ex. A ¶	Exhibit A to Declaration of Sanjay Krishnan, Ph.D. in Support of OpenAI's Opposition to Class Plaintiffs' Motion for Summary Judgment	2180
Lafferty ¶	Declaration of John D. Lafferty, Ph.D. in Support of Microsoft's Motions for Summary Judgment	1723

Msft Books MSJ	Defendant Microsoft Corporation's Memorandum of Law in Support Of Motion For Summary Judgment In Books Plaintiffs' Consolidated Cases	1692
Msft Books Reply SUMF ¶	Microsoft's Reply to Books Plaintiffs' Response to Rule 56.1 Statement of Undisputed Material Facts in Support of Microsoft's Motion for Summary Judgment	
Msft Books SUMF ¶	Rule 56.1 Statement of Undisputed Material Facts in Support of Microsoft's Books Motion for Summary Judgment	1694
Msft Books Opp.	Defendant Microsoft Corporation's Opposition To Books Plaintiffs' Motion For Partial Summary Judgment	2179
Msft News MSJ	Defendant Microsoft Corporation's Memorandum of Law In Support Of Motion For Summary Judgment In News Plaintiffs' Consolidated Cases	1699
Msft News Opp.	Defendant Microsoft Corporation's Opposition To News Plaintiffs' Motion For Partial Summary Judgment	2225
Msft News Reply	Defendant Microsoft Corporation's Reply Memorandum of Law In Support Of Motion For Summary Judgment In News Plaintiffs' Consolidated Cases	
Msft News SUMF ¶	Defendant Microsoft Corporation's Rule 56.1 Statement of Undisputed Material Facts in Support of Motion for Summary Judgment in News Plaintiffs' Consolidated Cases	1702
Msft Resp-BP Counter-SUMF ¶	Microsoft's Response to Books Plaintiffs' Counterstatement of Material Facts in Opposition to Microsoft's Motion for Summary Judgment	
NP Opp.	News Plaintiffs' Opposition to Defendants' Motions for Summary Judgment	2119
OAI Books Counter-SUMF ¶	Defendant OpenAI's Additional Facts Pursuant to Local Rule 56.1(b)	2237
OAI Books MSJ	OpenAI's Memorandum of Points and Authorities in Support of Motion for Summary Judgment	1718

OAI SUMF ¶	OpenAI's Statement of Undisputed Material Facts in Support of Its Motion for Summary Judgment	1735
Pls' Resp-OAI SUMF ¶	Class and News Plaintiffs' Response to OpenAI's Statement of Undisputed Material Facts in Support of Its Motion for Summary Judgment	2202

All statutory cites are to Title 17 unless otherwise indicated. All ECF cites are to MDL No. 1:25-md-3143-SHS-OTW unless otherwise indicated.

INTRODUCTION

What is most telling about Books Plaintiffs' briefs is how little they have to say about training LLMs, even though they brought this case purporting to challenge this groundbreaking technology. In the face of an overwhelming fair use defense, Books Plaintiffs devote most of their energy to splitting off ancillary claims (not raised in their original Complaint) based on discrete downloads or transfers of data—which they predictably spin as nefarious side deals. But these made-up claims do not change the answer to the central question in the case Books Plaintiffs brought: LLM training is fair use based on undisputed facts and established law.

The reason Books Plaintiffs strain to avoid the central issue in their own case is that this technology is an astounding new way of using written works. Advances in machine learning now allow engineers to use millions of ordinary written works to derive statistical patterns, from which new natural language text can be generated. These capabilities enable countless different types of uses. Users worldwide are now regularly experimenting with the technology to improve how they learn, live, and work. LLMs are a remarkable milestone in human ingenuity, one that harnesses the language and learnings in humanity's collected works to multiply what society can accomplish. Books Plaintiffs cannot dispute it, so they try to avoid it.

This Court should reject Books Plaintiffs' attempt to slice up the various activities that go into LLM development. Those activities include, of course, the literal steps that yield a finished model. But they also include research, testing, refinement, and so forth—much as developing a pharmaceutical drug entails R&D, trials, and other steps beyond mere production of the final chemical compound. The record in this case establishes that every use of Plaintiffs' works was reasonably necessary to LLM

development. If Plaintiffs are allowed to sever off critical aspects of that process to sustain separate claims, no innovator can safely develop new LLM technologies or pursue the next breakthrough.

This Court should also resist attempts by both sets of Plaintiffs to infuse the fair use analysis with predictions and policy preferences. From reckless claims that someone could use an LLM to start a war with China to questions about how LLMs might ultimately affect the future of life and work, none of it has any place in a litigated copyright dispute. What matters here is the evidence. And what that evidence shows beyond dispute is that the use of written works to train an LLM is fair use. Microsoft's motion should be granted.

ARGUMENT

I. Training An LLM Is Fair Use As A Matter Of Law.

Microsoft has made its fair use case at length, while refuting Plaintiffs' fundamental misunderstandings of the law and record. *See* Msft Books MSJ 9-33; Msft Books Opp. 9-35; Msft News MSJ 10-13; Msft News Opp. 7-13. Microsoft addresses below Books Plaintiffs' remaining fair use arguments (in both of their oppositions) as to the determinative first and fourth factors, as well as overlapping arguments made in News Plaintiffs' opposition as to LLM training. *Infra* § A (factor one arguments); § B (factor four arguments); *see also* Msft Books MSJ 24 (discussing second and third factors); Msft Books Opp. 24 (same). There is no avoiding the plain truth that LLM training is immensely transformative and has not hurt the markets for Plaintiffs' works.

A. Plaintiffs fail to undermine the transformative purpose and character of LLM training (Factor one).

Plaintiffs consume pages arguing that the use of books and news articles, tokenized and combined with trillions more tokens, for purposes of engineering a natural-language-speaking general-purpose artificial intelligence model is somehow "not 'transformative' in the copyright sense." NP Opp. 11. It would be surprising if copyright law were so senseless, and it is not.

LLM training is a new and different way to use books and articles. Plaintiffs do not dispute the basic fact that LLM training "is a new and different way to use" written works. Msft Books MSJ 11. The purposes for which authors write books and news articles and the ways readers consume them are not in dispute. Pls' Resp-OAI SUMF ¶¶ 95, 97; Msft Books Reply SUMF ¶¶ 18, 21. The record also details the fundamentally different way engineers use written works for LLM training—from tokenizing to using a transformer to deriving parameters to the vast, finished mathematical function that is an LLM. Msft Books Reply SUMF ¶¶ 3, 4.b-4.d, 12, 19; Pls' Resp-OAI SUMF ¶¶ 38, 42. Books Plaintiffs suggest that there are "disputed factual questions about the role of Plaintiffs' books in the training process," BP OAI Opp. 27, but there is no disputing that this role is profoundly different from the usual role for written works. Msft Books Reply SUMF ¶¶ 18-19.

It is really that simple. A use that "serves a new and different function" or has a "further purpose or different character" is transformative. Msft Books MSJ 10 (quoting *Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87, 96 (2d Cir. 2014); *Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith*, 598 U.S. 508, 528-29 (2023)). That inquiry is the heart of the fair use analysis because copyright law reserves to the public the right to

make inventive new uses of copyrighted works. The property interest enjoyed by authors—while powerful in its domain—is confined to traditional and reasonably likely uses, and protected only by copyright's bundle of exclusive rights. Books and News Plaintiffs have no right to prohibit (or be paid for) uses beyond that.

Generalization, not memorization, is the purpose of LLM training. Plaintiffs cannot meaningfully contest that the object of training an LLM using written works is to generalize—to "draw out, then sharpen, the relationships that reflect the structure of human language and knowledge, including semantics, grammar, facts, ideas, associations, and so forth." Msft Books MSJ 12 (citing Msft Books SUMF ¶¶ 4.d-4.e, 13); Msft Books Reply SUMF ¶¶ 1-2, 9-10, 14. That is what enables LLMs to respond helpfully to a vast array of unseen prompts with direct, accurate, and contextually apt strings of generated text. Msft Books SUMF ¶ 9; BP Resp-Msft SUMF ¶ 9 (admitting "the training objective for LLMs is next-token prediction" with "generalization" as a "hoped-for outcome"); *id.* ¶ 12. And there is likewise no meaningful dispute that every written work used in training is used to inflect the statistical parameters that endow the LLM with the ability to generalize. Msft Books Opp. 9-12; BP Resp-Msft SUMF ¶ 12; Msft Books Reply SUMF ¶¶ 10-12.

This is all borne out by the many undisputed uses of LLMs Microsoft and OpenAI have catalogued. Msft Books MSJ 3, 13; Msft Books Opp. 5-6, 10; OAI Books MSJ 1, 4-5. Plaintiffs do not and could never dispute that the GPT models do indeed succeed in countless uses across countless contexts. BP Resp-Msft SUMF ¶ 3; Pls' Resp-OAI SUMF ¶¶ 17, 19, 21. It is also not genuinely disputed that LLMs could not succeed at any of these uses unless training abstracted the structure of language and knowledge up and

away from underlying expression. Msft Books Reply SUMF ¶¶ 9, 12-13, 19-20. Generalization is why LLMs can do what they do.

What Books Plaintiffs call "hotly contested material facts" surrounding training, BP OAI Opp. 22, amount to nothing more than "different *characterizations* of the facts" and "would not affect the ultimate decision on the fair use question even if resolved in plaintiff[s'] favor." *Maxtone-Graham v. Burtchaell*, 803 F.2d 1253, 1259 (2d Cir. 1986) (emphasis added). Books Plaintiffs emphasize the concepts of "memorization" and "compression," suggesting that if anyone describes any part of the training process in those terms, that must mean that the secret true purpose of LLM training is to create a regurgitation machine. BP OAI Opp. 22-25. But Books Plaintiffs' own expert agrees that "memorization and generalization are not opposites," and the expert testimony Plaintiffs themselves cite explains that unintended memorization at one stage can give way to generalization in a finished model. Msft Books Reply SUMF ¶¶ 10, 13. The material question is whether the purpose of using works in the manner in which OpenAI used them was to create a model that generalizes, or merely to create a glorified database for distributing books and articles to users.

If Plaintiffs are insinuating the latter, they run directly into a mountain of undisputed evidence about the massive investment that OpenAI and Microsoft needed to make to develop the LLMs at issue. Surely those billions of dollars were not meant to make book-and-article vending machines. In the end, Plaintiffs' argument boils down to the same "regurgitation" analysis by their expert, Dr. Shan, who fed a GPT model passages of existing works in a lab setting and occasionally coaxed the model to return 30-word continuations of text. BP OAI Opp. 23. As Microsoft has explained, this adversarial prompting was so rarely successful that it actually *proves* how well the GPT

models generalize. Msft Books MSJ 22; Msft Books Reply SUMF ¶¶ 14, 15.a-15.d; Pls' Resp-OAI SUMF ¶¶ 86-89. That a model is theoretically *capable* of emitting text matching a work in the training set in rare circumstances in no way creates a fact dispute as to whether generalization is the intended *purpose* of LLM training. *See* Lafferty ¶¶ 276-78, 282 ("The relevant empirical question is whether the trained model behaves as though it retained the Asserted Works verbatim, not whether model capacity makes some amount of memorization theoretically possible."); Msft Books Reply SUMF ¶ 14.

For the same reason, Books Plaintiffs get no further with quotes from OpenAI engineers acknowledging that regurgitation is an inherent possibility of the training process. *See* BP OAI Opp. 23-24. What Plaintiffs leave out is that these engineers were intently focused on minimizing it, because that is part of training an LLM. *See* Msft Books Reply SUMF ¶ 14; Defs' Books Counter-SUMF ¶ 548; OAI SUMF ¶¶ 65-81. Plaintiffs do not genuinely dispute that OpenAI's training and deployment techniques are explicitly aimed at promoting generalization and preventing regurgitation. Msft Books Reply SUMF ¶¶ 13-14; Pls' Resp-OAI SUMF ¶¶ 66, 74-75, 78-79. Perhaps if a model developer *failed* at this, a plaintiff could claim that the model at issue is just a regurgitation machine. *See* Msft Books SUMF ¶¶ 4.e, 13; Lafferty ¶¶ 71, 74-76. But the vast range of undisputed uses for the GPT models—not to mention Dr. Shan's abject failure to raise an inference of real-world regurgitation—proves that they generalize in purpose and practice.

LLM training extracts unprotected elements like linguistic patterns, writing style, and factual information. Plaintiffs also cannot dispute that LLM training extracts unprotected elements of written works. They advance the

metaphysical critique that LLMs do not "learn" linguistic patterns or facts in the human sense. BP OAI Opp. 16, 22. For example, News Plaintiffs say that "LLMs don't 'learn facts'" because they "[redacted]" and instead just model the "frequency and arrangement of words." NP Opp. 9-10. This is irrelevant; neither Microsoft nor OpenAI claims that LLMs learn exactly like humans do. LLMs are trained by what engineers commonly call "machine learning" techniques to discern relationships reflecting generalizable facts, language, ideas, and so forth. The important point is not the *mode* of this learning but its *object*—to abstract away from the particular manner of expression to unprotectable elements across a written dataset. Msft Books Reply SUMF ¶¶ 3, 12, 19-20.

At times, Plaintiffs seem to resist altogether the notion that their works contain unprotectable elements—as though copying to access *any* aspect of their works is automatically unlawful. News Plaintiffs say, for example, that "[c]opying factual expression isn't transformative." NP Opp. 10. First off, LLM training is not bare "copying"—it is an undisputed technical reality that training maps statistical relationships through a model's parameters. Msft Books Reply SUMF ¶¶ 4.c-4.d, 12, 19-20. But even leaving that aside, Plaintiffs are wrong as a matter of law in claiming that their copyright extends to things like "information obtained via costly newsgathering," NP Opp. 10, or "patterns of language," BP OAI Opp. 22. The only thing in their works that has "intrinsic expressive value," NP Opp. 10, is the *expression*. Section 102(b) excludes information, methods, and concepts from the subject matter of copyright altogether.

Plaintiffs' argument is directly at odds with established law. As *Authors Guild v. Google, Inc.* explained, an author's "copyright does not extend to the facts

communicated by" their works. 804 F.3d 202, 224 (2d Cir. 2015). That is why *Authors Guild* found that copying entire books and then revealing a snippet of text with factual content to a user is fair use. *Id.* For much the same reason, the Copyright Act preempts state-law claims (apart from "narrow 'hot news'" claims) predicated on unauthorized use of uncopyrightable factual information, because such claims would otherwise improperly "expand the perimeters of copyright protection." *Nat'l Basketball Ass'n v. Motorola, Inc.*, 105 F.3d 841, 851-52 (2d Cir. 1997). This Court enforced that limit in dismissing News Plaintiffs' common law unfair competition claims based on the allegation that Defendants' products "produce informative text of the same general type" as news content. *N.Y. Times Co. v. Microsoft Corp.*, 777 F. Supp. 3d 283, 319-23 (S.D.N.Y. 2025). The same underlying principle bars Plaintiffs from claiming rights in unprotectable elements as part of the fair use analysis.

It makes no legal difference that Plaintiffs' works were not each strictly necessary to create the GPT models. Books Plaintiffs again argue that OpenAI had no "cognizable justification" for its use of their works because those works (versus other works authored by someone else) were not strictly *necessary* to create an LLM. BP OAI Opp. 25-26; *see also* NP Opp. 16-17, 19-20. As Microsoft's opposition explained (at 21), the use of written works to train an LLM is justified by its transformative purpose.

Warhol states plainly that "a use that has a distinct purpose is justified because it furthers the goal of copyright." 598 U.S. at 531. This is why copyright law reserves to the public the right to make new and different uses. Where a use is *not* transformative—that is, "where an original work and copying use share the same or highly similar purposes"—then a defendant may need an "independent justification," like the sort of

strict necessity Plaintiffs insist upon. *Warhol*, 598 U.S. at 532-33; *Whyte Monkee Prods., LLC v. Netflix, Inc.*, 174 F.4th 761, 794 n.11 (10th Cir. 2026) (under *Warhol*, only non-transformative uses require an "independent justification for … copying … which reasonably necessitated targeting"). For transformative uses, however, transformative purpose *is* the justification, "even *without* an independent (narrower) justification for copying." *Whyte Monkee*, 174 F.4th at 791 (emphasis added); *Warhol,* 598 U.S. at 531. The only remaining question is under the third factor, where "[c]omplete unchanged copying has repeatedly been found *justified* as fair use when the copying was *reasonably appropriate* to achieve the copier's transformative purpose and was done in such a manner that it did not offer a competing substitute for the original." *Authors Guild*, 804 F.3d at 221 (emphasis added).

The use of Plaintiffs' works to train an LLM bears that nexus to the transformative purpose of training. There is no material dispute that to train the GPT models, OpenAI needed so much training material that it had to use millions of works. BP Resp-Msft SUMF ¶ 5 (failing to dispute that LLMs of the scale of the larger accused models required "trillions of tokens" and admitting that even "[s]maller LLMs" were "trained on 300 billion tokens"); Msft Books Reply SUMF ¶¶ 5-6, 10-11, 28.b. Nor can Plaintiffs dispute that, as a technical matter, copying works is the only way to train an LLM—that is, the copies are not some gratuitous use unrelated to training. Msft Books Reply SUMF ¶¶ 4, 4.a-4.e, 7, 19; Pls' Resp-OAI SUMF ¶¶ 36-42; Lafferty ¶ 103; Callison-Burch Opp., Ex. B ¶¶ 59, 62.a. So while no *particular* author's work was *strictly* necessary, the works OpenAI selected and the manner in which it used them were "*reasonably appropriate*" for LLM training. *Authors Guild*, 804 F.3d at 221 (emphasis added). That is all the law requires.

Plaintiffs cannot distinguish seminal fair use cases. Plaintiffs also fail to distinguish the key fair use precedents that control this case. Msft Books MSJ 13-15 (citing *Authors Guild*; *Google LLC v. Oracle Am., Inc.*, 593 U.S. 1 (2021); *Sega Enters. Ltd. v. Accolade, Inc.*, 977 F.2d 1510 (9th Cir. 1992); and *Sony Comput. Ent., Inc. v. Connectix Corp.*, 203 F.3d 596 (9th Cir. 2000)).

Books Plaintiffs claim that all of these cases stand for one "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." BP OAI Opp. 28. This proposed test is hard to understand, let alone apply, and it is very much *not* the standard for transformativeness. A use is "transformative" when it "adds something new, with a further purpose," providing something "new and different from the original or expand[ing] its utility." *Authors Guild*, 804 F.3d at 214. Plaintiffs are running from that well-established standard because they cannot dispute that LLM training, by extracting unprotectable elements, is a new and further use beyond merely recapitulating the protected, creative expression in the underlying training set. *Supra* 6-8. That "clearly discernible, well-recorded purpose" makes LLM training fair use, regardless of how "functional" the copyrighted material is. *Graham v. Prince*, 2023 WL 3383029, at *13 (S.D.N.Y. May 11, 2023); *contra* BP OAI Opp. 28.

Plaintiffs also misread the cases. Books Plaintiffs claim that the use in *Authors Guild* involved only the "functional, as opposed to creative or expressive, aspects of the works," and that the copying was "functionally necessary." BP OAI Opp. 29. Plaintiffs fail to explain these characterizations. This is what the case was about: Google made exact, unauthorized copies of 20 million books in order to (a) make the text fully

searchable, while (b) delivering to users actual, identical snippets of that expression. *See* Msft Books MSJ 14 (detailing the case). If these verbatim copies and displays of text involve only "functionally necessary" uses aimed at "functional … aspects" of works, then so too does LLM training—which involves making full copies for purposes of deriving non-expressive elements of works in order to provide users a tool with new, further, and different functionality. *Supra* 9.

As for *Google v. Oracle*, Plaintiffs would rather not talk about it. They cannot dispute Microsoft's description of the case—that Google used verbatim copyrighted code for that code's ordinary purpose, but prevailed because the ultimate purpose was to "expand the use and usefulness" of products by offering a "highly creative and innovative tool," ultimately facilitating "new products." Msft Books MSJ 14 (quoting 593 U.S. at 30). News Plaintiffs protest that not every "innovative" use making an "invaluable contribution" to society is transformative. NP Opp. 12 (internal quotation marks omitted). No one has suggested otherwise. But that hardly justifies ignoring the considerable—indeed, dispositive—weight *Oracle* gave to these factors in a case where the defendant even used the plaintiff's works "for the same reason" the plaintiff did. 593 U.S. at 28-30. Here, LLM training is a profoundly new and different use of written works that *also* accomplishes the ultimate ends *Oracle* championed.

Plaintiffs' other attempt to distinguish *Oracle*—along with the Ninth Circuit's intermediate copying cases—points to the fact that the defendants there needed to copy the particular code at issue to make an interoperable product. BP OAI Opp. 30. That is factually true, and perhaps important in those cases because the use of the code had the same purpose and character as the original. *Oracle*, 593 U.S. at 28-30. But here,

OpenAI's uses are sufficiently justified by their immense transformativeness and therefore need no separate necessity justification. *Supra* 3-6.

Ross *does not apply to generative AI.* As Microsoft has explained, Msft Books Opp. 11-12, Plaintiffs' repeated citations to *Thomson Reuters Enterprise Centre GmbH v. Ross Intelligence Inc.*, 193 F.4th 271 (3d Cir. 2026), get them nowhere. *See, e.g.*, BP OAI Opp. 2, 5, 18-19, 21, 26. Books Plaintiffs must concede that "*Ross* is not a case about generative artificial intelligence," BP OAI Opp. 21, because *Ross* itself admonished that "generative AI" is different in its ability to generate new text, 193 F.4th at 284 n.7. Books Plaintiffs also concede that, unlike here, the parties in *Ross* "used the material at issue to sell similar end products to consumers." BP OAI Opp. 19. It is only through that lens that *Ross* held that the "intermediate step of using the headnotes to train an AI program …. *for the benefit of its legal-research platform*" was "minimally transformative." 193 F.4th at 284 (emphasis added). *Ross* has no bearing here, where OpenAI used books to train general-purpose LLM tools that bear no resemblance to traditional books. *Supra* 3-6.

B. Books Plaintiffs fail to create a dispute on substitutive harm in traditional markets (Factor four).

Microsoft's motion showed that Books Plaintiffs "are selling just as many books as they would have had ChatGPT and Copilot never been released," Msft Books MSJ 9, that "[t]he book authors who brought this suit confirmed the absence of traditional market harm," *id.* at 26, and that "analysis of sales of their works … show[s] no drop that could be attributable to the advent of Defendants' LLMs," *id.*; *accord* Msft Books Opp. 25. In their nearly 25,000 words of opposition briefing, Books Plaintiffs fail to meaningfully address these arguments, and their Rule 56.1 responses identify no

evidence tying lost sales to training the GPT models. *See* Msft Books Reply SUMF ¶¶ 22, 25. There is therefore no genuine dispute on market substitution. (We address the absence of harm to news markets at Msft News Reply 3-4.)

Books Plaintiffs respond by rehashing two legally deficient theories: (1) lost licensing revenue for the very transformative uses OpenAI made in training LLMs; and (2) theoretical competition from a claimed "flood" of AI-generated works. The arguments fail for the same reasons Microsoft has already addressed at length (at Msft Books MSJ 26-32; Msft Books Opp. 26-35) and below.

Books Plaintiffs cannot claim entitlement to licensing revenue from the transformative uses at issue. Books Plaintiffs claim they are entitled to licensing revenue for use of their works for LLM training, but fail to address the wall of contrary authorities Microsoft cited in its motion (at 27-28). That includes *Oracle*, which recognized and rejected reliance on "the theoretical market for licensing the very use at bar," 593 U.S. at 38 (citation omitted); *HathiTrust*, which rejected a "'lost sale' theory" based on the defendant's copies because "economic 'harm' caused by transformative uses does not count," 755 F.3d at 99-100; *Castle Rock Ent., Inc. v. Carol Publ'g Grp., Inc.*, which explained that "copyright owners may not preempt exploitation of transformative markets … by actually developing or licensing others to develop those markets," 150 F.3d 132, 145 n.11 (2d Cir. 1998); not to mention both *Bartz* and *Kadrey*, which reject reliance on a market for training data as circular.

Books Plaintiffs claim that *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 590-93 (1994), "require[s] consideration of existing or 'likely to develop markets,' even where the use may be transformative." BP OAI Opp. 35. It does not. *Campbell*, which involved a parodic rap song, held merely that harms to a derivative market for *non-*

parodic—and therefore non-transformative—rap songs are cognizable. 510 U.S. at 593. Not only that, *Campbell* hastened to add that harms flowing from "[t]he fact that a parody may impair the market for derivative uses by the very effectiveness of its critical commentary"—that is, flowing from the transformative use itself—are *not* "relevant." *Id.* So the relevant passage in *Campbell* not only fails to support Books Plaintiffs' market-for-training-data argument; it *also* undermines Books Plaintiffs' *market dilution* argument by rejecting the relevance of non-substitutive harms flowing from transformative uses.

Market dilution is not a legally cognizable theory, and Books Plaintiffs establish no dilution anyway. Microsoft has explained at length why Books Plaintiffs' market dilution theory fails as a matter of law and fact. Msft Books MSJ 30-32; Msft Books Opp. 28-33. Books Plaintiffs invoke *Ross* as somehow supporting the theory, BP OAI Opp. 37, but that case was not about indirect competition—"ROSS copied the headnotes to create a competing legal-research platform *in the same market* as Westlaw," 193 F.4th at 288 (emphasis added). Ultimately, though, there is no need to rehash this debate, because there is no need for this Court to resolve it. Even if market dilution were legally cognizable (it is not), Books Plaintiffs fail to create a fact dispute that they have "lost sales from AI-generated books," BP OAI Opp. 34.

Books Plaintiffs' only new argument on that score consists of quotes like this: "AI will do and is doing harm to me as an author by making available my words and ideas and style and discoveries …. I believe that it harms and will harm my ability to write and sell books." BP OAI Opp. 34 (quoting James Shapiro and citing other authors' similar quotes). Leaving aside the copyright-dystopian notion that "ideas" and "style" are an

author's private property, no author identifies any non-infringing AI-generated book, explains how that book would compete with theirs, or establishes that any AI-generated book would not exist but for the use of a Plaintiff's books in training. Msft Books Reply SUMF ¶¶ 21-22, 25, 26.a-26.d. Plaintiffs' self-serving testimony "consist[ing] largely of conclusory statements …, unsupported by documentary or other concrete evidence …, is simply not enough to create a genuine issue of fact in light of the evidence to the contrary." *Argus Inc. v. Eastman Kodak Co.*, 801 F.2d 38, 45 (2d Cir. 1986).

The use of written works to train LLMs is highly transformative and does not hurt Plaintiffs economically. This is quintessential fair use.

II. Microsoft Is Entitled To Summary Judgment On Plaintiffs' Distribution Claims.

Realizing that their attack on LLM training is unlikely to work, Plaintiffs manufacture direct infringement claims against Microsoft related to three distributions of training data: (1) Project Taxi, which entailed distribution of data from Microsoft's Bing Index to OpenAI for "computational analysis" related to LLM training; (2) OpenAI's transfer to Microsoft of GPT-3 training data to allow Microsoft to evaluate GPT-3 and associated technology; and (3) Project Mango, which entailed crawling and distribution to OpenAI of data used for LLM training. BP Msft Opp. 13-14, 23. Books Plaintiffs suggest that these should be viewed as unconnected to the larger project of LLM development, and thus not subject to the same fair use analysis.

Undisputed evidence refutes this characterization. Each of the distributions was an intermediate use directly related to development of LLMs—whether the literal steps of training or surrounding research and testing steps. *Infra* § A. As Microsoft has

several times explained,* all of these uses are fair for the same reasons as all others in the process—they advance the ultimate transformative purpose of creating LLMs, and these engineering copies are not market substitutes for Books Plaintiffs' works in the traditional books market. *Infra* § B; *see* Msft Books MSJ 23-24; Msft Books Opp. 9-23.

A. The distributions at issue were related to Defendants' collaboration in LLM training and development.

The publishers tried this same thing in *Authors Guild*. Apart from creation and operation of Google Books—the central uses that case was about—the publishers slivered off claims based on alleged "distributions" of digital copies to libraries, which the publishers "describe[d] … in more nefarious terms" than the facts could support. 804 F.3d at 228. Those claims failed. Because Plaintiffs' argument here is similarly based on describing ordinary conduct in nefarious terms, briefly detailed below are the undisputed facts material to the claims, and in particular fair use.

Project Taxi. Between 2022 and 2023, Microsoft transferred portions of its Bing Index to OpenAI. Msft Books Reply SUMF ¶ 16; Defs' Books Counter-SUMF ¶¶ 581-82, 593. This is the same index Microsoft has long maintained for purposes of its search engine (and, more recently, web grounding in Copilot). Msft Books SUMF ¶ 16; Msft News SUMF ¶ 70; Gaggara ¶¶ 7-8, 17. It is undisputed that OpenAI wanted portions of the Bing Index to conduct "computational analysis." Defs' Books Counter-SUMF ¶ 595; BP SUMF ¶ 582. As Padma Gaggara, Microsoft's corporate designee, explained in testimony emphasized by Plaintiffs, the data was not to be used for training

* Books Plaintiffs claim Microsoft did not move for summary judgment on one of the three distributions—Microsoft's downloading and transfer to OpenAI of training data, i.e., Project Mango. BP Msft Opp. 34. This is wrong. Microsoft sought summary judgment on all claims on fair use grounds, Msft Books MSJ 3, explicitly describing data acquisition and arguing that it is fair use, *id.* at 7, 10, 18.

a *commercial* model, but "to experiment [with] whether [the] data is useful for LLM training," Hurst Ex. 350, 101:16-102:7, and "to evaluate if such data … would … help improve the performance of the model," Hurst Ex. 351, 30:18-31:13; *see also* Msft Books MSJ 23; Msft Books Reply SUMF ¶ 17; Msft Resp-BP Counter-SUMF ¶¶ 29-32.

Books Plaintiffs' Rule 56.1 Statement asserts that only two of the Asserted Works had some text overlap with the Bing Index data. BP SUMF ¶ 940; Defs' Books Counter-SUMF ¶ 919; Hurst Ex. 352, 302:2-307:7. There is also no evidence that OpenAI specifically sought books in the Bing Index to use for computational analysis, let alone *Plaintiffs'* books. Msft Resp-BP Counter-SUMF ¶¶ 29, 31.

Seeking to depict the transfer as a sale, Books Plaintiffs note that OpenAI paid Microsoft in connection with Project Taxi. But the evidence is clear and undisputed that this was [redacted] Hurst Ex. 353, 57:8-25 (Jan. 21, 2026); Msft Resp-BP Counter-SUMF ¶¶ 33-34. That is, OpenAI was not paying for *data*; [redacted] which Microsoft engineers are required to do "even internally" for purely intra-Microsoft data storage. Hurst Ex. 353, 57:8-25.

Access to GPT-3 data. In 2021, OpenAI gave Microsoft access to portions of the already-tokenized data used to train the GPT-3 model on two occasions. Msft Books Opp. 22; Defs' Books Counter-SUMF ¶¶ 571, 573, 577; Msft Books Reply SUMF ¶ 30. One transfer was to enable Microsoft—which had contractual rights to the GPT-3 model once completed—[redacted]. Defs' Books Counter-SUMF ¶¶ 573-74, 578; Msft Resp-BP Counter-SUMF ¶¶ 44, 54. The other transfer (which did not include LibGen) was to enable Microsoft [redacted]. OAI Books Counter-SUMF ¶¶ 949-52; BP

SUMF ¶¶ 578-79. Like Project Taxi, there was nothing books-specific, let alone *Plaintiffs'*-books-specific, about Microsoft's need for access to GPT-3 training data. BP Msft Opp. 14, 23; Defs' Books Counter-SUMF ¶¶ 571, 578.

Attempting to conjure something looking like a transaction, Plaintiffs emphasize deposition testimony where one Microsoft witness described Project Taxi and GPT-3 training data access as "horse trading," to suggest Microsoft received "consideration for the distribution of Plaintiffs' Works." BP Msft Opp. 10, 37. Neither transaction had anything to do with *acquiring books* at all. As for the phrase "horse trading," providing access to large amounts of data is a significant technical and logistical undertaking. Both sides wanted access to data the other had to advance their joint development project, so in this instance it made sense for each side to facilitate access for the other. *See* Hurst Ex. 354, 184:3-185:10. That is a routine aspect of any high-tech collaboration between two independent companies, not some side-deal where Microsoft and OpenAI were trading Plaintiffs' books with one another. Msft Resp-BP Counter-SUMF ¶ 41.

Project Mango. As for Project Mango in 2023, there is no dispute that it was a joint effort by Microsoft and OpenAI to obtain training data that OpenAI could then use for training (in its discretion). Books Plaintiffs admit this purpose. BP SUMF ¶ 601. For its part, Microsoft crawled URLs of domains that OpenAI identified, streaming the data directly to OpenAI. Defs' Books Counter-SUMF ¶¶ 599-601, 604-06. As with Project Taxi, OpenAI's payments to Microsoft in connection with Project Mango [redacted]. Hurst Ex. 355 at -014; Defs' Books Counter-SUMF ¶ 600; Msft Resp-BP Counter-SUMF ¶ 42. Although Dr. Shan purports to have identified URLs associated with 27 Asserted Works that match the domains that were to be crawled by Microsoft, Krishnan Ex. A ¶ 56, he did not identify any content from LibGen or other

shadow library domains in the data crawled by Microsoft. Defs' Books Counter-SUMF ¶ 604; Msft Books Reply SUMF ¶¶ 31-32.

Plaintiffs make much of a Microsoft employee's description of Project Mango as a "secret." They use this to claim it was a "secret deal" or "'secret' workaround." BP Msft Opp. 13, 33, 36. But there is nothing untoward about Microsoft limiting information regarding a project with a particular, strategically important collaborator to need-to-know employees—it only underscores that Microsoft's purpose was not a broad data sale, but a discrete initiative tied to training. Defs' Books Counter-SUMF ¶ 606; Msft Resp-BP Counter-SUMF ¶ 40.

B. Defendants' transfers of data to facilitate LLM training are fair use.

The intermediate, engineering-focused data transfers described above are fair use as a matter of law. Each was a "reasonable … step towards training an LLM," undertaken in the context of a longstanding collaboration that existed for the purpose of developing LLMs and LLM-based technology. *Kadrey v. Meta Platforms, Inc.*, 788 F. Supp. 3d 1026, 1048 (N.D. Cal. 2025). And as to each, the determinative first and fourth factors weigh heavily in favor of fair use. (The second- and third-factor analyses are the same here as for all other uses. *See* Msft Books MSJ 24; Msft Books Opp. 24.)

1. The transfers were for the transformative purpose of LLM training.

a. Training an LLM is a highly complex technological endeavor. At the most basic level, the steps for training an LLM entail data acquisition, dataset creation, tokenizing, pre-training, and post-training. Msft Books SUMF ¶¶ 4, 4.a-4.e. But engineers working on LLM development must also conduct extensive research, evaluation, refinement, and so forth. Lafferty ¶¶ 67-68, 71, 169; Callison-Burch Opp.,

Ex. B ¶ 62.d. And because the essential inputs for LLMs are computerized data, all of the many tasks engineers perform in developing LLMs may require making an internal copy. Lafferty ¶¶ 69, 103; Callison-Burch Opp., Ex. B ¶¶ 59-64. As explained above (at 16-19), all of the transfers at issue here were of data for either literal training itself (like Project Mango) or for evaluating and testing LLMs (like Project Taxi or GPT-3 data access). And because LLM training is profoundly transformative, the first factor weighs as strongly in favor of fair use for these transfers as for all other training-focused uses.

Plaintiffs' attempts to detach these uses from LLM training are based on two shallow tricks. The first is a word game involving different senses of the term "LLM training." For example, with respect to Project Taxi, Books Plaintiffs point to testimony from Microsoft employee Gaggara that OpenAI was "not supposed to use [the Bing Index] for training," and instead that the data was for "experiment[s]." BP Msft Opp. 27 (quoting Hurst Ex. 350, 128:6-15). *Aha!*, Books Plaintiffs declare—this proves that Project Taxi was not for training LLMs and is therefore "unrelated to any allegedly transformative use." BP Msft Opp. 26. But the "experiment[s]"—the "computational analysis"—OpenAI did were *undisputedly* for purposes of determining "whether [the] data is useful for LLM training," Hurst Ex. 350, 101:16-102:7; *supra* 16-17, including by using it to train "smaller model[s]" for experimentation, Hurst Ex. 350, 48:20-49:24, 94:24-95:2; Msft Resp-BP Counter-SUMF ¶ 29. The *only* purpose of the use was therefore in furtherance of LLM training. And Books Plaintiffs do not even try to articulate some other, non-LLM-related purpose for any of the transfers at issue.

The second trick is to depict Project Taxi and Project Mango as using Plaintiffs' "Works … as bargaining chips to extract money and data." BP Msft Opp. 23. The implication of this strange locution seems to be that Microsoft was *selling Plaintiffs'*

works to OpenAI, but there is no evidence of this. OpenAI needed large amounts of data for LLM training, and Microsoft, as a development partner, helped with data acquisition or access on two occasions. For Projects Taxi and Mango, OpenAI's payments [redacted], precisely as you would expect for two collaborators on a larger project. *Supra* 17-18. And though OpenAI did later give Microsoft access to portions of the GPT-3 training data (the second transfer discussed above (at 17-18)), the transfer did not include LibGen and was for experimenting with GPT-3.

b. Plaintiffs also make the legal argument that even if the purpose of Project Taxi and Project Mango were OpenAI's training of LLMs, "Microsoft [cannot] prevail on a fair use defense by latching on to the conduct of OpenAI, a third-party recipient." BP Msft Opp. 28.

Authors Guild forecloses the argument. It dealt with "Google's provision of digitized copies [of books] to libraries" so that the libraries could "use their digital copies to enable the very kinds of searches that" the court had already found to be fair use. 804 F.3d at 228-29. The court explained that "[i]f the library had created its own digital copy to enable its provision of fair use digital searches, the making of the digital copy would not have been infringement." *Id.* at 229. It did not "become an infringement" because the libraries "contracted with Google" so Google could "use its expertise and resources" to do the same. *Id.*

The same logic applies here. OpenAI typically sourced training data itself for the fair-use purpose of training LLMs; that Microsoft assisted with this in two discrete instances does not change the legal analysis—particularly in a context where two longtime collaborators were working towards the same transformative and mutually beneficial end. Defs' Books Counter-SUMF ¶ 599; Msft Resp-BP Counter-SUMF ¶ 41.

Just as it would have been fair use for OpenAI to collect the data itself for training, *supra* 2-15, it was fair use for Microsoft to collect the data for OpenAI to use to train.

Books Plaintiffs' cases purportedly holding that "the conduct of third-party recipients of copyrighted works is irrelevant to the *distributor's* fair use defense" do not apply here. *See* BP Msft Opp. 28, 36. Their lead cases, typified by *American Geophysical Union v. Texaco Inc.*, hold merely that a single defendant cannot claim a use is fair merely because it is for a "company doing research." 60 F.3d 913, 924 (2d Cir. 1994). That is nothing like this case, where all uses are new and different ones undertaken for the purpose of training LLMs. *See* Msft Books Opp. 15-16 (refuting applicability of *Texaco*). Just as far afield are Books Plaintiffs' citations to cases about "copy shops" that allow anyone to make copies of any copyrighted work, for profit and for the same purpose as the original. BP Msft Opp. 28-29. Books Plaintiffs do not even attempt to explain how Microsoft's discrete data transfers to a longtime collaborator resemble a public copy shop.

2. Data transfers between partners in LLM development do not substitute in traditional markets for copyrighted works.

Factor four's substitution analysis also weighs just as heavily in favor of fair use with respect to the data transfers as it does for all other uses. Indeed, the lack of market impact from the few discrete data transfers at issue is even more obvious. First, these are transfers between two longstanding technological collaborators meant to facilitate engineering steps for LLM training projects already in flight. That is not a traditional market into which Books or News Plaintiffs sell or license their works. And second, these transfers involved large, varied, and unique datasets of which Plaintiffs' works, if any, would have amounted to an infinitesimal fraction. For example, the point of

Project Taxi was to provide OpenAI with portions of the Bing Index—large chunks of the internet—specifically for computational analysis; the traditional market for books could not logically have satisfied that demand.

Books Plaintiffs ignore the surrounding context of these transfers and treat them as bare sales of copies of their works. BP Msft Opp. 31-32. They ask the Court to "infer[]" harm to the market for their "ebooks," "audiobooks," and "works in connection with Amazon Alexa." BP Msft Opp. 31, 33, 37. But that inference is illogical and therefore cannot suffice as evidence of harm. *Cf. Hachette Book Grp., Inc. v. Internet Archive*, 115 F.4th 163, 193 (2d Cir. 2024) (plaintiff is only entitled to "*logical* inferences *where appropriate* in assessing the fourth … factor") (emphasis added). Again, OpenAI could not have gone to the audiobook market to get web data from the Bing Index for experimenting with LLM training. And Microsoft could not have assembled stray ebooks to replicate GPT-3-specific training data [redacted] with GPT-3. These transformative uses therefore could not plausibly have substituted for any sale in any market Books or News Plaintiffs traditionally exploit. *See Authors Guild*, 804 F.3d at 224-25 (factor four weighed in favor of fair use where "it would be a rare case" that the transformative use "could provide a significant substitute for the purchase of the author's book").

Along with the first factor, the fourth factor decisively weighs in favor of fair use as a matter of law.

III. Microsoft Cannot Be Held Liable For OpenAI's Downloading From Shadow Libraries.

Books Plaintiffs' last attempt to sliver off a separate claim is based on OpenAI's downloading from shadow libraries. But this claim could never be maintained against Microsoft under any theory.

First, Microsoft has explained several times why OpenAI's downloading from shadow libraries is fair use. The ultimate purpose of the downloads was the same transformative purpose as all other challenged uses in this case: training LLMs. Msft Books MSJ 17-20; Msft Books Opp. 13-16. OpenAI did not download works as a common dorm-room pirate, to stock a central digital lending library, or for any other ordinary reason anyone has previously used books. It did so for a transformative purpose far outside of the book markets Plaintiffs seek to exploit.

Second, even if OpenAI's downloads were not fair use as a matter of law, Microsoft cannot be held liable for them. As Microsoft has explained, undisputed evidence establishes that Microsoft played no role in OpenAI's downloading of training content from LibGen or any other shadow library. Msft Books MSJ 18; Msft Books Opp. 13, 35; Msft Books Reply SUMF ¶¶ 30-32. Books Plaintiffs do not dispute that Microsoft cannot be held directly liable for this conduct, so summary judgment should be granted with respect to any direct infringement claim based on OpenAI's downloads from shadow libraries.

And *third*, Microsoft cannot be held secondarily liable for this conduct. As explained in Microsoft's separate motion for judgment on the pleadings, Books Plaintiffs' only theory of contributory liability was abolished in *Cox Communications, Inc. v. Sony Music Entertainment*, 607 U.S. 583 (2026).

That leaves only Books Plaintiffs' sweeping suggestion that Microsoft is vicariously liable for "OpenAI's direct infringements." BP Msft Opp. 39. As explained in Microsoft's opposition (at 36-42), this attempt to categorically hold Microsoft vicariously liable for every act of a fully independent company, over a multiyear period, is utterly unprecedented. It is also impossible to reconcile with Microsoft and OpenAI's joint development and collaboration agreements (JDCAs), which provide in black and white that OpenAI retains all control over LLM training, including selection of datasets. Msft Books Opp. 38.

Also beyond any dispute is that vicarious liability could never attach for OpenAI's downloads from LibGen, all of which took place in 2021. Books Plaintiffs attempt to predicate Microsoft's purported control of OpenAI on a supposed "right and ability to inspect and 'retrain' OpenAI's models." BP Msft Opp. 39. But as previously explained, Msft Books Opp. 39, that clause was not included in a JDCA until *2023*, Hurst Ex. 356 at -590, years after OpenAI's downloads and training of GPT-3 and GPT-3.5 using that downloaded material, OAI SUMF ¶ 11; BP SUMF ¶¶ 519-21, 526. That leaves Plaintiffs with nothing but the bare fact that OpenAI "downloaded and stored" the data on an ordinary Microsoft cloud server, BP Msft Opp. 39—a server that Microsoft undisputedly had no visibility into or ability to modify, Msft Books Opp. 39-40.

Because there is no basis for holding Microsoft liable for OpenAI's downloading from shadow libraries, summary judgment should be granted on any such claims.

CONCLUSION

The Court should grant summary judgment in favor of Microsoft on all claims asserted against it.

Dated: November 20, 2026

Respectfully submitted,

/s/ Annette L. Hurst

ORRICK, HERRINGTON & SUTCLIFFE LLP
Annette L. Hurst (admitted pro hac vice)
Daniel Justice (admitted pro hac vice)
Robert S. Shwarts
Catherine Lui (admitted pro hac vice)
Amanda H. Schwartz (admitted pro hac vice)
The Orrick Building
405 Howard Street
San Francisco, CA 94105-2669
Telephone: (415) 773-5700
Facsimile: (415) 773-5759
ahurst@orrick.com
djustice@orrick.com
aschwartz@orrick.com

Christopher J. Cariello
Lisa T. Simpson
Marc Shapiro
Vanessa Sorrentino
Hilda Obeng
51 West 52nd Street
New York, NY 10019
Telephone: (212) 506-3778
Facsimile: (212) 506-5151
ccariello@orrick.com
lsimpson@orrick.com
mrshapiro@orrick.com
vsorrentino@orrick.com
hobeng@orrick.com

Sheryl Koval Garko (admitted pro hac vice)
Laura Brooks Najemy (admitted pro hac vice)
222 Berkeley Street
Boston, MA 02116
Telephone: (617) 880-1800
Facsimile: (617) 880-1801
sgarko@orrick.com
lnajemy@orrick.com

Isaac Sawlih Behnawa (admitted pro hac vice)
Johnathan Vaknin (admitted pro hac vice)
355 S. Grand Ave. Suite 2700
Los Angeles, CA 90071
Telephone: (213) 612-2179
Facsimile: (213) 612-2499
ibehnawa@orrick.com
jvaknin@orrick.com

Monica Svetoslavov (admitted pro hac vice)
Louis Andrew Schreiber
2100 Pennsylvania Avenue NW
Washington, DC 20037
Telephone: (202) 339-8400
Facsimile: (202) 339-8500
msvetoslavov@orrick.com
aschreiber@orrick.com

FAEGRE DRINKER BIDDLE & REATH LLP
Jeffrey S. Jacobson
1177 Avenue of the Americas
New York, New York 10036
Telephone: (212) 248-3191
jeffrey.jacobson@faegredrinker.com

Jared B. Briant (admitted pro hac vice)
Kirstin L. Stoll-DeBell (admitted pro hac vice)
Timothy Scull (admitted pro hac vice)
Shelby Pickar-Dennis (admitted pro hac vice)
1144 15th Street, Suite 3400
Denver, Colorado 80202
Telephone: (303) 607-3588
jared.briant@faegredrinker.com
kirstin.stolldebell@faegredrinker.com
timothy.scull@faegredrinker.com
shelby.pickar-dennis@faegredrinker.com

Lora A. Brzezynski (admitted pro hac vice)
Brianna Lynn Silverstein (admitted pro hac vice)
Faith Maggard (admitted pro hac vice)
1500 K Street NW, Suite 1100
Washington, DC 20005
Telephone: (202) 230-5000
Facsimile: (202) 842-8465
lora.brzezynski@faegredrinker.com

brianna.silverstein@faegredrinker.com
faith.maggard@faegredrinker.com

Carrie A. Beyer (admitted pro hac vice)
320 South Canal Street, Suite 3300
Chicago, IL 60606-5707
Telephone: (312) 569-1000
Facsimile: (312) 569-3000
carrie.beyer@faegredrinker.com

Elizabeth M.C. Scheibel (admitted pro hac vice)
2200 Wells Fargo Center, 90 S. 7th Street
Minneapolis, MN 55402
Telephone: (612) 766-7000
Facsimile: (612) 766-1600
elizabeth.scheibel@faegredrinker.com

Andrew Lawrence Van Houter
600 Campus Drive
Florham Park, NJ 07932
Telephone: (973) 549-7018
Facsimile: (973) 360-9831
andrew.vanhouter@faegredrinker.com

Attorneys for Defendant Microsoft Corporation

CERTIFICATE OF WORD COUNT COMPLIANCE

Pursuant to Local Civil Rule 7.1(c), the foregoing Reply complies with the Court's order of August 7, 2026 (ECF 1664), as it contains 7,483 words, according to the word processing program used to prepare it.

/s/ Annette L. Hurst
Annette L. Hurst
Attorney for Defendant
Microsoft Corporation