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 Opposition To Books Plaintiffs' Motion For Partial Summary Judgment

TABLE OF CONTENTS

Page

TABLE OF AUTHORITIES iii
TABLE OF CITATION CONVENTIONS vi
INTRODUCTION 1
COUNTER-STATEMENT OF MATERIAL FACTS 4
Defendants developed LLMs to create a foundational technology with a broad range of uses. 4
Predictions that LLM-based products will undermine the market for books have not materialized. 6
ARGUMENT 9
I. Training An LLM Is Fair Use As A Matter Of Law. 9
A. All challenged uses are for the transformative purpose of training an LLM (Factor one). 9
1. Plaintiffs ignore undisputed material facts establishing the transformativeness of LLM training. 9
2. Plaintiffs fail to undermine the transformativeness of any challenged use. 12
a. OpenAI's acquisition from "shadow libraries" was for a transformative purpose. 13
b. LLM training is profoundly transformative. 17
c. Transfer of data between Defendants was for training-related uses. 22
3. The commercial nature of the uses does not weigh against fair use. 23
B. The nature of the underlying works and the amount used carry little weight in light of the transformativeness of LLM training (Factors two and three). 24
C. Training an LLM does not cause substantial market harm in any cognizable market (Factor four). 24
1. Undisputed material facts establish the absence of cognizable harm in any traditional book market. 24
2. Plaintiffs identify no cognizable market harm from any challenged use. 26
a. Plaintiffs' claims of harm from downloading or transfer of training data are empty. 26
b. Plaintiffs' market dilution theory is legally untenable and factually speculative. 28

c. There is no legally cognizable market for training data. 33
II. Plaintiffs Fail To Establish Microsoft's Copying Or Distribution As To The Vast Majority Of Direct Infringement Claims. 35
III. Books Plaintiffs Are Not Entitled To Summary Judgment That Microsoft Is Vicariously Liable For OpenAI's Direct Infringement. 36
A. Plaintiffs fail to show that Microsoft had the legal right and practical ability to supervise OpenAI's data acquisition and training activities. 37
B. Plaintiffs fail to show that Microsoft derived a direct financial benefit from OpenAI's alleged infringement of their works. 40
CONCLUSION 42
CERTIFICATE OF WORD COUNT COMPLIANCE 46

TABLE OF AUTHORITIES

Page(s)

Cases

A&M Records, Inc. v. Napster, Inc.,
239 F.3d 1004 (9th Cir. 2001) 14

Adobe Sys. Inc. v. Canus Prods., Inc.,
173 F. Supp. 2d 1044 (C.D. Cal. 2001) 37

American Geophysical Union v. Texaco Inc.,
60 F.3d 913 (2d Cir. 1994) 15, 24

Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith,
598 U.S. 508 (2023) 10, 12, 14, 19, 21, 27

Arica Inst., Inc. v. Palmer,
970 F.2d 1067 (2d Cir. 1992) 30

Authors Guild, Inc. v. HathiTrust,
755 F.3d 87 (2d Cir. 2014) 25, 33

Authors Guild v. Google, Inc.,
804 F.3d 202 (2d Cir. 2015) 11, 15, 21, 23, 24, 25, 26, 27, 31

Banff Ltd. v. Ltd., Inc.,
869 F. Supp. 1103 (S.D.N.Y. 1994) 37

Bartz v. Anthropic PBC,
787 F. Supp. 3d 1007 (N.D. Cal. 2025) 9, 11, 16, 33

Cambridge Univ. Press v. Patton,
769 F.3d 1232 (11th Cir. 2014) 34

Cartoon Network LP, LLLP v. CSC Holdings, Inc.,
536 F.3d 121 (2d Cir. 2008) 13, 35

Castle Rock Ent., Inc. v. Carol Publ'g Grp., Inc.,
150 F.3d 132 (2d Cir. 1998) 33

Cengage Learning Inc. v. Google LLC,
786 F. Supp. 3d 611 (S.D.N.Y. 2025) 41

Consumers Union of U.S. v. Gen. Signal Corp.,
724 F.2d 1044 (2d Cir. 1983) 23, 30

In re Demetriades,
58 F.4th 37 (2d Cir. 2023) 36

Eldred v. Ashcroft,
537 U.S. 186 (2003) 18

Ellison v. Robertson,
357 F.3d 1072 (9th Cir. 2004) 41

EMI Christian Music Grp., Inc. v. MP3tunes, LLC,
844 F.3d 79 (2d Cir. 2016) 40

Flo & Eddie, Inc. v. Sirius XM Radio, Inc.,
821 F.3d 265 (2d Cir. 2016) 12, 13, 14

Folsom v. Marsh,
9 F. Cas. 342 (C.C. D. Mass. 1841) 30

In re Google Generative AI Copyright Litig.,
809 F. Supp. 3d 903 (N.D. Cal. 2025) 40

Google LLC v. Oracle America, Inc.,
593 U.S. 1 (2021) 11, 19, 20, 25, 29

Hachette Book Grp., Inc. v. Internet Archive,
115 F.4th 163 (2d Cir. 2024) 14, 20, 27

Kadrey v. Meta Platforms, Inc.,
788 F. Supp. 3d 1026 (N.D. Cal. 2025). 10, 11, 12, 16, 27, 28, 29, 31, 32, 33

Luvdarts, LLC v. AT&T Mobility, LLC,
710 F.3d 1068 (9th Cir. 2013) 40

Michael Grecco Prods., Inc. v. Valuewalk, LLC,
345 F. Supp. 3d 482 (S.D.N.Y. 2018) 41

NXIVM Corp. v. Ross Inst.,
364 F.3d 471 (2d Cir. 2004) 30, 32

On Davis v. The Gap, Inc.,
246 F.3d 152 (2d Cir. 2001) 29

Oracle Am., Inc. v. Google LLC,
886 F.3d 1179 (Fed. Cir. 2018) 19

Perfect 10, Inc. v. Amazon.com, Inc.,
508 F.3d 1146 (9th Cir. 2007) 37

Perfect 10, Inc. v. Giganews, Inc.,
847 F.3d 657 (9th Cir. 2017) 41

Romanova v. Amilus Inc.
138 F.4th 104 (2d Cir. 2025) 21

Sega Enterprises Ltd. v. Accolade, Inc.,
977 F.2d 1510 (9th Cir. 1992) 11, 20, 30, 31

Shapiro, Bernstein & Co. v. H.L. Green Co.,
316 F.2d 304 (2d Cir. 1963) 37

Softel, Inc. v. Dragon Med. & Sci. Commc'ns, Inc.,
118 F.3d 955 (2d Cir. 1997) 37

Sony Comput. Ent., Inc. v. Connectix Corp.,
203 F.3d 596 (9th Cir. 2000) 24

Sony Music Ent. v. Cox Commc'ns, Inc.,
93 F.4th 222 (4th Cir. 2024) 41

Thomson Reuters Enter. Centre GmbH v. Ross Intelligence Inc.,
2026 WL 2932557 (3d Cir. Sept. 29, 2026) 11, 12

United States ex rel. Taylor v. GMI USA Corp.,
774 F. Supp. 3d 645 (S.D.N.Y. 2025) 36

Triangle Publ'ns, Inc. v. Knight-Ridder Newspapers, Inc.,
626 F.2d 1171 (5th Cir. 1980) 30

A.V. ex rel. Vanderhye v. iParadigms, LLC,
562 F.3d 630 (4th Cir. 2009) 15, 33

VHT, Inc. v. Zillow Grp., Inc.,
918 F.3d 723 (9th Cir. 2019) 40

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
BP MSJ	Class Plaintiffs' Memorandum of Law in Support of Motion for Partial Summary Judgment
BP SUMF ¶	Class Plaintiffs' Corrected Rule 56.1 Statement of Undisputed Material Facts in Support of Class Plaintiffs' Motion for Partial Summary Judgment
Compl. ¶	Consolidated Class Action Complaint, ECF 183
Defs' Books Counter-SUMF ¶	Defendants' Response to Books Plaintiffs' Rule 56.1 Statement
Lafferty ¶	Declaration of John D. Lafferty, Ph.D. in Support of Microsoft's Motions for Summary Judgment
Msft Books MSJ	Defendant Microsoft Corporation's Memorandum of Law in Support Of Motion For Summary Judgment In Books' Plaintiffs' Consolidated Cases
Msft Books SUMF ¶	Rule 56.1 Statement of Undisputed Material Facts in Support of Microsoft's Books Motion for Summary Judgment
Msft News Opp.	Defendant Microsoft Corporation's Opposition To News Plaintiffs' Motion For Partial Summary Judgment
NP MSJ	News Plaintiffs' Combined Summary Judgment Brief
NP SUMF ¶	News Plaintiffs' Corrected Combined Rule 56.1 Statement of Undisputed Material Facts
OAI SUMF ¶	OpenAI's Statement of Undisputed Material Facts in Support of Its Motion for Summary Judgment
Simpson Ex.	Declaration of Lisa T. Simpson in Support of Defendant Microsoft's Oppositions to the Books and News Plaintiffs' Motions for Partial Summary Judgment

Tucker ¶	Declaration of Catherine Tucker, Ph.D. in Support of Defendant Microsoft Corporation's Motion for Summary Judgment in Books Plaintiffs' Consolidated Cases
Usdan ¶	Declaration of Jordan Usdan in Support of Microsoft Corporation's Motions for Summary Judgment

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

It may well be that among the countless things LLMs are useful for, one is to help writers work on books—the way several Books Plaintiffs themselves use the technology. But that does not make *generating books* the purpose of building an LLM any more than the purpose of manufacturing a hammer is to generate chairs. Like a hammer, an LLM is a general-purpose tool that empowers humans to pursue whatever *their* ends may be. And rarely (if ever) has a new technology been so widely adopted, so avidly deployed, for so many beneficial uses in such a short time. Yet somehow Plaintiffs' two motions for summary judgment and the thousands of supposedly material facts they assert scarcely even acknowledge what LLMs are, how they are created, and how people are already using them. Plaintiffs can ignore the profoundly transformative purpose of LLM training all they want. It will not change the incontestable reality that the use of written works to train an LLM yields a foundational technology that democratizes learning and original creation—the very ends copyright law aims to promote.

It turns out that this democratization is exactly what Books Plaintiffs fear. They devote their entire motion to the possibility that because LLMs are helpful for writing, people can (and some likely do) use them to help create new books. In their view, this means that the very *purpose* of training an LLM must be to generate these new books, and thereby supplant the famed authors who brought this case. And they claim that the use of their works to create a new technology that (among countless other things) allows others to compete with them—even through original, non-infringing works that the public values—cannot be fair use. This copyright-upside-down theory is deeply flawed.

To begin with, Books Plaintiffs' extreme closeup on one use misses the point of an LLM. Undisputed evidence establishes that LLMs are not built to do one thing, but

many—organization, research, data processing, brainstorming, software coding, and so on. Defendants did not set out to develop LLMs to accomplish any particular use themselves, but to bring *users* a tool with broad functionality that helps them accomplish whatever they wish. And undisputed evidence establishes that the transformative use of books (among other works) to train LLMs serves this broad range of uses. In short, an LLM is not a book machine. It is an everything machine.

Books Plaintiffs also fundamentally misunderstand copyright law as shielding them from creative competition. For two centuries, the law has protected fair uses precisely *because* they spur more creative output. It has repeatedly approved transformative uses that yield innovative products for public enjoyment and edification, condemning only uses that usurp an author's exclusive rights. This conclusion does not change because a group of famous authors fear, without evidence, that a transformative use might threaten their dominance. Nor does copyright law endorse Books Plaintiffs' implicit value judgment that the use of LLM technology to write more quickly or consider more drafts makes the resulting works less worthy. Copyright law wisely leaves that for the public to decide, rather than entrenching dominant authors or their preferred methods of generating creative output.

In any event, Plaintiffs' theory of market harm has another problem: It hasn't happened. Hard economic evidence shows that LLM technology has had no impact on Books Plaintiffs' bottom lines. That is why instead of evidence, Books Plaintiffs sensationalize with sound bites from other prognosticators, including some who work in academic research functions at Microsoft or OpenAI. But predictions do not become facts by repeating them over and over. Ultimately, all Plaintiffs point to as evidence for market dilution is the acknowledgement of some self-publishers on Amazon that their

publications contain *some* amount of AI-generated text, translation, or even cover art. Plaintiffs offer no proof that any of these works on Amazon—which include "low-content" books like journals and planners—compete with their own. After years of public use of LLMs, the notion that using *The Pelican Brief* to train the GPT models results in new books that people buy instead of *The Pelican Brief* is just a fanciful hypothetical contradicted by the real economic evidence.

So Books Plaintiffs retreat to another approach: use the word "piracy" a lot. Their lead claim is that regardless of whether LLM training itself is fair use, OpenAI's downloading of training data from "pirate libraries" on the internet is independently infringing because it is "[l]ike teenagers in the 2000s" downloading music. But this ignores the crux of the fair use analysis—*the purpose* of the downloading. Napster users downloaded music in order to listen for free. OpenAI downloaded books not to read them or distribute copies to others, but to tokenize them, pool the tokens with trillions more from various sources, and create a wondrous AI that can talk (and code, and conquer advanced math, and decipher ancient texts, etc.). That transformative purpose makes all the difference as a matter of copyright law and common sense. Books Plaintiffs cannot defeat fair use by ignoring the very heart of the fair use inquiry.

The parties do agree on one thing: The important fair use questions here should be decided on summary judgment as a matter of law. But only Microsoft and OpenAI faithfully apply the law to the record. As two courts have already held, using books to train LLMs is profoundly transformative. And as the Supreme Court, Second Circuit, and other courts have made clear, groundbreaking uses like this cause no harm to any market that copyright holders are entitled to exploit. Because LLM training is fair use, Books Plaintiffs' motion should be denied.

COUNTER-STATEMENT OF MATERIAL FACTS

Books Plaintiffs' final Rule 56.1 statement is 157 pages long and contains 940 purported material facts. Though Defendants must respond to each asserted fact in their counter-statement, the below discussion is confined to the only two issues that matter: (1) the transformative purpose of developing LLMs; and (2) the absence of any harm to the market for Books Plaintiffs' works. To avoid unhelpful duplication, Microsoft also incorporates background from its other summary judgment briefs where indicated.

Defendants developed LLMs to create a foundational technology with a broad range of uses.

The basics of what an LLM is and how it is trained are amply described in Defendants' briefing. *See* Msft Books MSJ 6-8; OAI Books MSJ 5-9. LLMs are AI models capable of receiving an input (a prompt) and generating output (a response to the prompt) by predicting the next token in a sequence of text. Msft Books SUMF ¶¶ 1, 9. They are trained on extraordinary amounts of text sourced from a wide variety of written content. *Id.* ¶¶ 4, 7-8; OAI SUMF ¶¶ 4, 8, 10. The statistical relationships between the trillions of tokens in the training set, which the LLM gleans during the training process, are embedded as model weights, which dictate the LLM's predicted output. Msft Books SUMF ¶¶ 4(c)-(d), 12. The LLM can then be built into products or tools that people can put to a vast range of uses. *Id.* ¶ 3.

Books Plaintiffs want to talk about only one thing. Ignoring the countless uses for LLM technology, they point to evidence that OpenAI was aware that LLMs could be "excellent at writing" and "very very good at writing stories." BP MSJ 13, 26 (citing BP SUMF ¶¶ 709-17). It is true that people find LLM-based tools valuable for all kinds of

writing tasks, from real estate listings to meeting summaries to presentation notes, and so on. It is also true that Defendants foresaw these use cases (among many others), that they worked to create LLMs that write well, and that OpenAI considered books as training material because books would be one of many sources of tokens enabling these capabilities. Msft Books SUMF ¶ 7; OAI SUMF ¶¶ 9-10, 136; Defs' Books Counter-SUMF ¶¶ 706-19. What the evidence does not support, however, is Books Plaintiffs' impossible logical leap: The claim that because LLMs are *good* at writing stories, "OpenAI developed its LLMs with *the purpose of creating competing books*," BP MSJ 28 (emphasis added), as though OpenAI wanted to create competing books so it decided to train an LLM.

This does not follow. What Books Plaintiffs' narrow frame misses is that LLMs are, and are intended to be, a *general-purpose technology* that can do many things: "[A]utomate the process of tax forms," Simpson Ex. 296, 354:22-355:2; "summarize the … consultations between patients and doctors," Simpson Ex. 297, 116:19-117:6; "understand [how] proceeds are distributed" from a "bond sale," Simpson Ex. 298, 74:13-24; "explain legal concepts in lay terms," Simpson Ex. 320, 66:17-68:24; find "recipes and stain removal tips," Simpson Ex. 342, 30:10-17. All of these uses depend on an LLM's natural language processing capabilities. And those capabilities depend on an immense and varied corpus of training data, including books. Msft Books SUMF ¶¶ 7-9; OAI SUMF ¶ 9; Lafferty ¶¶ 69, 90, 97.

Not only are LLMs inherently general-purpose technology, Defendants themselves envisioned their respective work in developing the GPT models as ultimately yielding *general-purpose tools*. OpenAI—who selected and acquired the training data and trained the GPT models, Msft Books SUMF ¶¶ 30-31—understood that LLMs would

be used in products that accomplish a broad range of tasks. The "vision" for ChatGPT, one of OpenAI's LLM-based products, was to "1. Answer any question I have 2. Aid me with my cognitive work 3. Perform[] personal tasks 4. Commercially transact on my behalf." Simpson Ex. 299 at -421.

Similarly, when Microsoft partnered with OpenAI—supplying infrastructure and investment in exchange for access to finished GPT models, Defs' Books Counter-SUMF ¶¶ 121-22, 630, 637-38, 670-74—its aim was to develop tools that can accomplish "professional tasks across various domains," Simpson Ex. 300 at -368—from "meeting recap in 10 languages," Simpson Ex. 301 at -039, to "enabl[ing] non-technical customers to easily access their data and metrics via natural speech," Simpson Ex. 302 at -722, to "vertical domain specific [natural language processing] offerings (e.g., Medical, Finance, HR) across all [Microsoft] Office features," *id.* And Microsoft thought its LLM-powered Copilot products could help people "Prepare for a pet … Make a meal … Plan a vacation … Create a sports image … Decorate an apartment." Simpson Ex. 300 at -367.

In short, the purpose of using books to train LLMs is not to accomplish *any one* of the things described above. It was to develop a foundational technology that would enable billions of people to do, build, and create *all* of them and more.

Predictions that LLM-based products will undermine the market for books have not materialized.

None of those products have displaced books. Books Plaintiffs have been claiming that LLMs will "endanger authors' ability to make a living" for years now. Compl. ¶ 3. Economic evidence shows they are plain wrong. The introduction of the GPT models has had no impact on Books Plaintiffs' sales. Msft Books SUMF ¶¶ 22-24; Defs' Books Counter-SUMF ¶¶ 877-88.

Unable to dispute this, Books Plaintiffs rely heavily on soundbites from Dr. Brent Hecht, a researcher at Microsoft who is concurrently an Associate Professor of computer science at Northwestern University. News Plaintiffs also rely heavily on Hecht soundbites, since their predictions that LLMs will destroy journalism have also failed to come true. *See* NP MSJ 1, 2, 11, 13, 53, 54, 69, 72, 73, 74. They point to Hecht statements like his "doom loop" hypothesis, where he suggests that content creators should be paid because if they are not, they may stop creating the training data needed to develop LLMs. BP MSJ 15. Or they note Hecht's warning that LLM developers will "be[] accused of the 'largest theft of labor in human history.'" NP SUMF ¶ 1652. Given the leading-man treatment Plaintiffs give Hecht in their briefing, a quick biography is in order.

Dr. Hecht works in a role in which he "explore[s] new research directions." Simpson Ex. 303, 117:17-118:7; *see* Simpson Ex. 304, 18:4-16. As Microsoft's corporate designee put it, Hecht is a "kind of intellectual provocateur" who holds something like "an academic position within [the] company," with a tenure-like freedom to "speak [his] mind." Simpson Ex. 305, 85:16-86:1, 173:22-174:9. Or as Hecht himself sees it, a "key part of [his] job" is to "pose hypotheses" that he is "not sure about" and then "test those hypotheses." Simpson Ex. 304, 169:20-170:2.

He does not set Microsoft's generative AI strategy, direct the partnership with OpenAI, negotiate Microsoft's data access or display agreements, or set institutional priorities for Microsoft's decade-long responsible AI initiatives. *See* Usdan ¶¶ 39-40; Simpson Ex. 305, 247:1-6; Simpson Ex. 303, 117:17-118:7; Simpson Ex. 304, 185:16-22. He is not an expert in intellectual property law (or any other legal field), history, economics, book publishing, or journalism. *See* Simpson Ex. 304, 287:19-23; Usdan

¶ 39. Hecht himself testified that he does not speak for the company. Defs' Books Counter-SUMF ¶ 751. His colleagues agree. Usdan ¶ 39. And his name did not "[r]ing any bells" to Microsoft CEO Satya Nadella, nor did Nadella recall seeing Hecht's September 2023 presentation hypothesizing a "doom loop." Simpson Ex. 306, 137:13-138:20.

It is also important to place Hecht's words in their academic context, particularly the idea of "data dignity." This theory was advanced primarily by two other Microsoft researchers. Usdan ¶ 38. It posits that *all* data—social media musings, personal information, search history, etc.—result from human labor that should (theoretically) be compensated. *See id.*; Defs' Books Counter-SUMF ¶ 771; Simpson Ex. 307, 56:3-57:20. The theory is "not in any way … related to the current copyright regime." Simpson Ex. 307, 57:7-20. And Hecht's purported fear is that if data generation is not treated as compensable labor, the supply will dry up and impede the technology's wider adoption. *See* Defs' Books Counter-SUMF ¶¶ 764-65. It is thus Hecht's prescription for how to create sustainable AI. It is most certainly *not* a normative claim that *copyright law* protects John Grisham's labor from competition, nor that it guarantees Grisham compensation—a claim he concedes he would be entirely unqualified to make in any event. *See* Simpson Ex. 304, 287:19-23.

If all of that sounds far afield, that's because it is. So why do Plaintiffs devote so much airtime to an arcane academic data-as-labor hypothesis advanced by a "provocateur" who concededly does not speak for Microsoft? The answer is the only thing in this subsection that matters: They do it to distract from undisputed economic evidence that decisively refutes their predictions that LLM technology would harm the market for their works.

ARGUMENT

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

Microsoft's Summary Judgment Motions explained why LLM training is fair use as a matter of law. Msft Books MSJ 9-33; Msft News MSJ 10-13. The analysis is straightforward. Acquisition and use of books to train an LLM is vastly different from traditional uses of those books. And neither the training process nor the resulting LLM substitutes for traditional consumption of these books as expressive works. Highly transformative uses that do not affect existing markets are quintessentially fair.

A. All challenged uses are for the transformative purpose of training an LLM (Factor one).

1. Plaintiffs ignore undisputed material facts establishing the transformativeness of LLM training.

A party cannot win summary judgment by ignoring material facts. Yet you could reach the end of Books Plaintiffs' 40-page motion and have little idea what an LLM is, let alone how and why written content is used to train one.

Plaintiffs devoutly avoid the details because to state them is to concede how "spectacularly" and "exceedingly" transformative it is to use a book to train an LLM. *Bartz v. Anthropic PBC*, 787 F. Supp. 3d 1007, 1019-22 (N.D. Cal. 2025). LLM training entails: (a) acquiring a large volume of text to assemble a training corpus; (b) "tokeniz[ing]" the text by breaking it into small pieces, then assigning a unique integer to each token; (c) embedding these tokens in high-dimensional mathematical space as vectors; and (d) iteratively passing data through transformer layers to enable a model to adjust its "parameters," which the model will ultimately use to predict the next token in a sequence. Msft Books SUMF ¶¶ 4, 12; OAI SUMF ¶¶ 36-44, 50. There is also no dispute that the technological objective of that process is "generalization," by which the

LLM maps unprotected semantic relationships and knowledge about the world, enabling it to respond to new, previously unseen input. Msft Books SUMF ¶ 9; OAI SUMF ¶ 38; *see* BP SUMF ¶¶ 502-03, 511. Nor can Books Plaintiffs deny that the ultimate purpose of training an LLM is to create an innovative new technology capable of fielding a limitless range of unseen queries with newly generated output. Msft Books SUMF ¶¶ 9-13; OAI SUMF ¶¶ 6, 48-49, 57-58.

Plaintiffs focus instead on one among countless *downstream uses* of LLMs—that people could use them to write new original books that might compete with theirs. BP MSJ 26; BP SUMF ¶¶ 740-44. If so, good for these new authors and good for the public, which gets more creative works to choose from. But there is no evidence that this particular *use* of LLM-based tools was Defendants' *purpose* in creating an LLM. *Supra* 4-6; *infra* 17-19. Fair use, after all, "is an *objective* inquiry into what [the defendant] does with an original work." *Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith*, 598 U.S. 508, 545 (2023) (emphasis added). Books Plaintiffs are just identifying one among a near-limitless range of *end-user* uses, from synthesizing information to developing lesson plans to proofing documents to searching the internet to preparing a work presentation and on and on. The purpose of LLM training is to create a technology that can do *all of those things* in *all of those* contexts, at the direction of a human user. Msft Books SUMF ¶¶ 9-13; OAI SUMF ¶¶ 20-22, 56, 133, 313-14.

Two courts have already examined the use of books to train an LLM and both found it transformative. *Kadrey v. Meta Platforms, Inc.* emphasized that LLM training "ingests text to learn 'statistical patterns' of how words are used together in different contexts," yielding "innovative tools that can be used to generate diverse text and

perform a wide range of functions" and that carry the "potential to exponentially multiply creative expression." 788 F. Supp. 3d 1026, 1044-45 (N.D. Cal. 2025). *Bartz* similarly recognized the aim of "iteratively map[ping] statistical relationships between every text-fragment and every sequence of text-fragments so that a completed LLM could receive new text inputs and return new text outputs" does not use a "given work's creative elements" but "grammar, composition, and style that the underlying LLM distilled from thousands of works." 787 F. Supp. 3d at 1021-22.

And numerous seminal decisions establish that when a use is as transformative as LLM training, that tilts not only the first factor, but all of them in favor of fair use. Msft Books MSJ 13-15. The Second Circuit's decision in *Authors Guild v. Google, Inc.*, 804 F.3d 202 (2d Cir. 2015); the Supreme Court's decision in *Google LLC v. Oracle America, Inc.*, 593 U.S. 1 (2021); and the Ninth Circuit's decision in *Sega Enterprises Ltd. v. Accolade, Inc.*, 977 F.2d 1510 (9th Cir. 1992), all strongly support that proposition. Of these, Books Plaintiffs address only *Authors Guild*, claiming it is "miles away" because the transformative use there was to "convey[] to the public information *about* books." BP MSJ 27. (The tool in *Authors Guild* also displayed snippets of books, but Books Plaintiffs do not mention that part. 804 F.3d at 208-10.) Books Plaintiffs' purported distinction is just an observation that *Authors Guild* addressed a different type of transformative use. If anything, LLM training is miles *ahead*—"among the most transformative many of us will see in our lifetimes." *Bartz*, 787 F. Supp. 3d at 1033.

Meanwhile, every case Books Plaintiffs rely upon is one in which the court found that the use in question was *not* transformative. For example, Books Plaintiffs cite *Thomson Reuters Enterprise Centre GmbH v. Ross Intelligence Inc.*, where the defendant used copyrighted Westlaw headnotes to create an AI legal research platform

that would compete with Westlaw. The Third Circuit recognized that the "lodestar" of the analysis is "whether the copier maintains a 'distinctive purpose.'" 2026 WL 2932557, at *6 (3d Cir. Sept. 29, 2026). It then held that copying and using Westlaw headnotes for the same purpose Westlaw does is "minimally transformative." *Id.* In the same breath, the Third Circuit distinguished "generative AI" because it can "generate original expression," noting *Bartz* and *Kadrey*'s holdings. *Id.* at *6 n.7. Books Plaintiffs' other favorite case, *Warhol*, is similarly just a case where a defendant made use of a copyrighted work for "substantially the same purpose." 598 U.S. at 526 (photorealistic version of a portrait of Prince was not transformative).

The only salient takeaway from these cases is that whether a use is transformative is indeed the all-important question in the fair use analysis. Plaintiffs cannot prevail by simply ignoring extensive and undisputed evidence resolving that question resoundingly in favor of fair use.

2. Plaintiffs fail to undermine the transformativeness of any challenged use.

Books Plaintiffs try to escape the transformativeness of LLM training by slicing up different training-focused uses, claiming each is subject to a different fair-use analysis. They break training into (a) acquisition of books, focusing on OpenAI's downloading from "pirate sites" (aka shadow libraries); (b) training itself; and (c) the transfer of training materials from one Defendant to another. Notably, *News* Plaintiffs eschew the slice-and-dice approach, a good indication that Books Plaintiffs' approach is artificial. But we can analyze the uses separately. It makes no difference because undisputed evidence establishes that each is for the same transformative purpose of training an LLM, so all must be considered "in light of [this] ultimate, highly

transformative purpose." *Kadrey*, 788 F. Supp. 3d at 1047-48; *see Flo & Eddie, Inc. v. Sirius XM Radio, Inc.*, 821 F.3d 265, 270 n.4 (2d Cir. 2016) (question is "whether the *ultimate* use … is permissible") (emphasis added).

a. OpenAI's acquisition from "shadow libraries" was for a transformative purpose.

First is Books Plaintiffs' claim based on "OpenAI's unauthorized acquisition of [Books Plaintiffs'] copyrighted books by downloading from LibGen and via downloading the Books3 dataset." BP MSJ 19. This fails for several reasons.

No downloading claim can be asserted against Microsoft. At the outset, Books Plaintiffs' account of OpenAI's downloading confirms that they cannot assert a direct infringement claim against Microsoft for any such downloads. Msft Books MSJ 17-18. Microsoft did not do the downloading at issue, nor engage in any other volitional act causing OpenAI's downloading. Msft Books SUMF ¶¶ 30-31; Defs' Books Counter-SUMF ¶¶ 156, 160, 180, 210. *See Cartoon Network LP, LLLP v. CSC Holdings, Inc.*, 536 F.3d 121, 131 (2d Cir. 2008) ("*Cablevision*") (volitional conduct required for direct infringement). Nor, as explained in greater detail below (at 36-42), can Microsoft be held vicariously liable because it had no right or ability to supervise OpenAI's data acquisition. Indeed, the most Books Plaintiffs even allege is that OpenAI chose to use LibGen to train more quickly "in order to secure Microsoft's funding." BP MSJ 5-6. So Microsoft cannot be held liable for the conduct.

OpenAI's downloading was solely for LLM training. The "downloading claim" fails in any event because it ignores the ultimate purpose of OpenAI's downloading. Books Plaintiffs assert that "[l]ike teenagers in the 2000s, OpenAI downloaded books for free while knowing it could have purchased or licensed copies

precisely to avoid paying for them." BP MSJ 20. But *all* fair uses involve someone knowingly engaging in use of a work without the owner's prior authorization. What matters is the *purpose* of that use, because "[t]he same copying may be fair when used for one purpose but not another." *Warhol*, 598 U.S. at 533. The problem with "teenagers in the 2000s" was that they downloaded songs from Napster for the purpose of listening to them rather than buying CDs, a traditional and non-transformative use. *A&M Records, Inc. v. Napster, Inc.*, 239 F.3d 1004, 1015-17 (9th Cir. 2001). The problem in Plaintiffs' other cited authority, *Hachette Book Group, Inc. v. Internet Archive*, was similar. The defendant there "distributed [digital books] to its users in full, for free" for those users to read the same way they otherwise would on their Kindle. 115 F.4th 163, 181 (2d Cir. 2024). Plaintiffs are just citing cases about non-transformative uses again.

OpenAI's downloading was not for a traditional purpose. It was for the profoundly transformative, previously unimaginable purposes outlined above (at 9-12). Msft Books SUMF ¶¶ 19-20; OAI SUMF ¶¶ 10-11, 104; Defs' Books Counter-SUMF ¶¶ 186, 227, 262-68, 402, 517-20. Only by ignoring that could one equate OpenAI's downloading with dorm-room piracy or a digital lending library.

Books Plaintiffs ask the Court to do just that, claiming it does not matter if OpenAI downloaded books "for a supposedly 'transformative' use." BP MSJ 20. The Court must ignore the purpose of OpenAI's downloads, they say, because it "must analyze each separate use separately." *Id.* at 21. But analyzing a use separately does not mean analyzing it myopically. Cases evaluating the purpose and character of an intermediate use in a longer process routinely focus on that use's ultimate purpose. *See Flo & Eddie*, 821 F.3d at 270 n.4 (noting the "fair-use analysis applicable to" an internal

database of copies "is bound up with whether the ultimate use of the internal copies is permissible"). For example, *Authors Guild* did not evaluate Google's unauthorized "scann[ing], render[ing] machine-readable, and index[ing] more than 20 million books" as isolated uses which are no different from an ordinary reader photocopying whole books at the library. 804 F.3d at 208. It evaluated those uses in light of the ultimate purpose of creating Google Books. *Id.* at 214-19. Similarly, *A.V. ex rel. Vanderhye v. iParadigms, LLC* held that "archiving" of student papers "had an entirely different function and purpose than the original works"—to create anti-plagiarism software—notwithstanding "the fact that there was no substantive alteration to the works." 562 F.3d 630, 639 (4th Cir. 2009).

Books Plaintiffs claim support from Second Circuit case law they say has "repeatedly rejected claims that a downstream use can be used to bless other acts of copying." BP MSJ 21. The cases they cite stand for nothing like that. Their lead case, *American Geophysical Union v. Texaco Inc.*, was about Texaco's "archival" photocopying of journal articles, where "the dominant purpose of the use [wa]s a systematic institutional policy of multiplying the available number of copies of pertinent copyrighted articles by circulating the journals among employed scientists … *thereby serving the same purpose* for which additional subscriptions are normally sold, or … for which photocopying licenses may be obtained." 60 F.3d 913, 924-25 (2d Cir. 1994) (emphasis added). All *Texaco* said was that this use does not become "'transformative' … simply" because it was "in the course of doing research," which of course is the original purpose of scientific journals in the first place. *Id.* In other words, the ultimate purpose in *Texaco* was not transformative at all. The "dominant purpose"

of OpenAI's downloading, by contrast, was for the transformative purpose of training LLMs.

Books Plaintiffs offer the glancing counter that because some of the books downloaded "were *never* used for LLM training" they must have been "used to stock OpenAI's central library," which all OpenAI employees could "access … for whatever purpose they wished." BP MSJ 22. There was no central library. Follow Books Plaintiffs' citations and the "central library" turns out to be basic electronic computer storage where OpenAI puts data collected for training. Defs' Books Counter-SUMF ¶¶ 244-60. That OpenAI ultimately elected not to use some of the books that it downloaded to train the models at issue does not mean that downloading that content was for some other purpose, let alone make OpenAI a librarian or archivist. *See id.* ¶¶ 242-43. Nor do Plaintiffs point to any evidence that OpenAI used the downloaded data for any non-training related purpose.

Once these arguments fall away, all that remains of Books Plaintiffs' argument is the musing in *Bartz* that perhaps obtaining materials "from a pirate site … is inherently, irredeemably infringing," even if those materials are used for LLM training. BP MSJ 21 (quoting 787 F. Supp. 3d at 1025). *Bartz* cites nothing to support this, and ultimately did "not decide th[e] case on" any such basis. 787 F. Supp. 3d at 1026. As Microsoft explained in its motion (at 17-18), neither the case law nor the logic underlying fair use supports that proposition. That is why *Kadrey* rightly rejected *Bartz*'s rumination, instead following § 107's directive that courts consider—not ignore—the ultimate purpose of the use. 788 F. Supp. 3d at 1047-48. This Court should do the same.

b. LLM training is profoundly transformative.

Books Plaintiffs similarly fail to diminish the transformativeness of LLM training itself.

Books Plaintiffs' theory is fundamentally flawed. Plaintiffs' principal argument is that LLM training is "nontransformative because it [1] exploited the protected expression in Plaintiffs' works [2] for the purpose of [3] developing a competing commercial product." BP MSJ 25. This single sentence conceals three demonstrably erroneous premises, denoted by the brackets in the quote.

First, the evidence uniformly shows that LLM training *is not* designed to "exploit … protected expression." Again, the undisputed purpose of training is to generalize up and away from any particular way of expressing ideas or concepts, instead encoding the fundamentals of language and relationships between facts about the world. *Supra* 9-12, 17-19; Msft Books MSJ 6-8, 12-13. The end goal is a complex mathematical function able to generate *new* expression. *Supra* 9-12, 17-19. Books and all other writing are vital to that process simply because *they contain expression*—that is, because they contain datapoints that contribute to the LLM's map of language and knowledge. *Supra* 5, 9-10, 17-18.

Books Plaintiffs emphasize that "a model's ability to generate 'book-like' text derives from … training on books." BP MSJ 27. True, LLM training uses books, along with many other types of written works, to endow an LLM with natural language capabilities. Msft Books SUMF ¶¶ 5, 7-8, 10, 20; OAI SUMF ¶¶ 4, 10; Defs' Books Counter-SUMF ¶¶ 502-08. Those capabilities are derived from the whole training set, including books, and can then be used to generate "book-like" text (however one defines the term). *Supra* 3-5, 18, 34. But the mere fact that the content used to train an LLM

contains protected expression does not remotely establish that the purpose of LLM training is to *exploit that protected expression*. After all, books (and other works) also contain unprotected ideas, concepts, facts, and so forth. Defs' Books Counter-SUMF ¶ 555. And "every idea, theory, and fact in a copyrighted work becomes instantly available for public exploitation at the moment of publication." *Eldred v. Ashcroft*, 537 U.S. 186, 219 (2003). If an LLM derives the usual length of chapters or story pacing from detective novels, for example, it has derived only unprotected ideas and methods, and in this way become better at generating "book-like text." Msft Books SUMF ¶¶ 19-20; OAI SUMF ¶¶ 58-60. That does not exploit protected expression.

Second, Books Plaintiffs offer no evidence that the *purpose*—that is, the object or goal—of using books to train an LLM is to create competing books. Again, the evidence shows that model trainers use books to train LLMs because books contain large amounts of tokens reflecting use of grammar, syntax, concepts, facts, and ideas about the world. Msft Books SUMF ¶¶ 5-6, 19-20; Defs' Books Counter-SUMF ¶¶ 453, 507, 790. Conversely, an LLM's ability to generate writing depends not just on books, but on everything else in the training dataset. Msft Books SUMF ¶ 20. This all leads back to the basic purpose of enabling a model to generalize and therefore generate original responses to a vast range of previously unseen prompts. *Supra* 4-6, 17-19.

Books Plaintiffs claim that the "record is replete with evidence that OpenAI developed its LLMs … to 'supplant' Plaintiffs in this case." BP MSJ 28. That misrepresents the record. Plaintiffs' evidence suggests merely that end-users might use LLMs for writing tasks. BP MSJ 26-28. For example, they point to statements that LLMs are "very very good at writing stories" and could "generate 'economic disruption.'" *Id.* at 26. Again, yes, it is true that someone could use an LLM to help them write a

book. But saying that generating competing books is the *purpose* of training an LLM is like saying that the purpose of the internet is retail sales because the internet is a "very very good" tool for selling things and "generate[s] economic disruption" at brick-and-mortar stores. That OpenAI was aware that LLMs are good for writing tasks does not mean that OpenAI's "objective" purpose in training the GPT models was to "supplant Plaintiffs in this case." *See Warhol*, 598 U.S. at 545; *supra* 9-11.

Third, in all events, Books Plaintiffs' argument rests on the legally erroneous premise that "developing a competing [] product" is a substitutive, and therefore non-transformative, use of a copyrighted work. BP MSJ 25. The idea appears to be rooted in *Kadrey*'s novel "market dilution" theory—that if the use of a work to train an LLM ultimately helps someone create a competing, non-infringing work, that competition counts as market harm under factor *four*. *Id.* at 25-29. But *Kadrey* does not suggest that this creative competition makes a use less *transformative* under factor *one*. Indeed, Books Plaintiffs' theory would swallow the transformativeness inquiry entirely: No matter how new and different LLM training is from ordinary use of books, if the resulting technology can be used by others to "mimic high-quality creative expression," BP MSJ 2, displace the "personal labor" of authors, *id.* at 26, or compete with book authors going forward, LLM training is not transformative. *See* Msft News Opp. 12-13 (addressing News Plaintiffs' version of this argument).

No case has ever held or even suggested this, and several refute it. In *Google v. Oracle*, for example, the Supreme Court reviewed the Federal Circuit's finding that the use of the defendant's code "in a competing platform" could not be fair use. 593 U.S. at 16 (quoting *Oracle Am., Inc. v. Google LLC*, 886 F.3d 1179, 1210 (Fed. Cir. 2018)). The Court rejected the idea that copyright imposes this broad restriction on competition.

The use was transformative, the Court held, because it "expand[ed] the use and usefulness" of smartphones, "offer[ed] programmers a highly creative and innovative tool," and facilitated "new products." *Id.* at 29-30. "'An attempt to monopolize the market by making it impossible for others to compete runs counter to the statutory purpose of promoting creative expression.'" *Id.* at 39 (quoting *Sega*, 977 F.2d at 1523-24).

Sega, which the Supreme Court quoted favorably in *Oracle*, is also illuminating. *See* Msft Books MSJ 15 (discussing *Sega*). There too the plaintiff claimed that "because [the defendant] copied [the plaintiff's] object code in order to produce a competing product," it could not be fair use. 977 F.2d at 1522. And there too the Ninth Circuit rejected the argument. The "ultimate purpose" was indeed "the release of" competing video games, but this weighed *in favor* of fair use: "It is precisely this growth in creative expression, based on the dissemination of other creative works and the unprotected ideas contained in those works, that the Copyright Act was intended to promote." *Id.* at 1522-23.

To be clear, this does not mean that any use of a work that somehow results in competition is automatically transformative. Non-transformative uses—like the "digital copies" in *Hachette* that merely "ma[de] [the] authors' works available to read," 115 F.4th at 181—are not automatically fair just because they increase competition. But where the use *is* transformative—as, for example, in *Oracle* or *Sega*—the fact that it yields competition from new, original creation is a *transformative good*, not a *substitutive evil*. And so the analysis circles back, as it always will, to the all-important question Plaintiffs are trying to avoid: Is the purpose and character of LLM training

transformative? Undisputed evidence establishes that it is, so the competition and innovation LLMs yield weighs heavily in favor of fair use, not against it.

LLM training is justified by its transformative purpose. Books Plaintiffs' secondary argument against LLM training is that "copying of Plaintiffs' works was not 'reasonably necessary' to achieve [a] purportedly transformative purpose." BP MSJ 29 (caps omitted). The argument is rooted in the concept that for a use to be fair, it must have a "justification." According to Plaintiffs, there is no such justification because "no individual work or category of works (such as books) was necessary for model performance." *Id.* at 29-30.

Books Plaintiffs cite *Warhol* to support their argument, but *Warhol* actually destroys it. *Warhol* explains that "[i]n a broad sense, 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." 598 U.S. at 531 (emphasis added). To support this principle, *Warhol* relies on none other than *Authors Guild*, in particular the Second Circuit's observation that "new, transformative" uses serve the aims of copyright law. *Warhol* thus makes clear that where a use is transformative, that "transformative purpose[]" *itself* "justifie[s]" the use. *Id.*; *see also Romanova v. Amilus Inc.*, 138 F.4th 104, 116-18 (2d Cir. 2025) (detailing case law finding sufficient justification where uses were transformative).

LLM training easily satisfies that test. Using a book to train an LLM is justified as a new, different, and highly transformative use. And the way books are used for LLM training—extracting unprotected elements to derive statistical patterns across a training set—is a "reasonably appropriate" use of those books. *Authors Guild*, 804 F.3d at 221; *see* Msft Books MSJ 24.

c. Transfer of data between Defendants was for training-related uses.

Books Plaintiffs last assert that "Defendants' distribution of Plaintiffs' works to each other and third parties was not fair use." BP MSJ 38 (caps omitted). News Plaintiffs make a similar version of the argument, calling the transfers between Defendants "horse trading" and insinuating that the data was used as "currency"—for what, exactly, they never say. NP MSJ 31. These arguments fail for the simple reason that any transfer of training data between Defendants was for purposes related to LLM training and evaluation, the same transformative purpose as all other challenged uses.

Undisputed evidence establishes this for each of the discrete transfers Books Plaintiffs note. First, on several occasions, Microsoft provided OpenAI with portions of its Bing search index that OpenAI used for research and experimentation into what sorts of training data would improve LLM performance. Msft Books SUMF ¶¶ 16-17; Defs' Books Counter-SUMF ¶¶ 581-82, 595-97. As explained in Microsoft's summary judgment motion (at 23), OpenAI used this for "computational analysis" to assess whether "web data can be useful" in training LLMs, not to train any accused model. Simpson Ex. 308 at -041; Simpson Ex. 309, 35:18-22, 101:16-102:18; Defs' Books Counter-SUMF ¶ 596.

Similarly, on two occasions, OpenAI gave Microsoft engineers temporary access to a portion of tokenized training data for GPT-3, which enabled Microsoft to test GPT-3's performance and experiment with how to incorporate the finished model into its own products. Defs' Books Counter-SUMF ¶¶ 573-75, 577-79; Simpson Ex. 310, 185:20-187:22; Simpson Ex. 312, 151:14-25; Simpson Ex. 347, 14:4-20, 20:15-25; Lafferty ¶¶ 67, 72. While Plaintiffs try to spin these transfers as some separate, purposeless currency

exchange, the undisputed evidence shows they were part and parcel of Defendants' partnership in LLM development and deployment. These uses, like all the others Plaintiffs challenge, were new and different ways to use books.

And "Project Mango" was likewise for purposes of LLM training. At OpenAI's direction, Microsoft used its Bing search crawling infrastructure to crawl a portion of the internet, streaming data directly to OpenAI. Defs' Books Counter-SUMF ¶¶ 599-601, 604-06. OpenAI undisputedly used the data for LLM training. *Id.* ¶ 601. Although OpenAI reimbursed Microsoft for the costs associated with its crawling services, that does not make OpenAI's use any different or any less transformative. *See* Defs' Books Counter-SUMF ¶ 600; *contra* BP MSJ 39.

3. The commercial nature of the uses does not weigh against fair use.

Though the challenged uses were generally of a commercial nature, Books Plaintiffs are wrong that this carries significant weight in the fair-use analysis. *See* BP MSJ 20, 32-33, 39. The flaw in their reasoning is the same one infecting all of their arguments: They ignore that LLM training is a phenomenally transformative use. The Second Circuit has "repeatedly rejected the contention that commercial motivation should outweigh a convincing transformative purpose and absence of significant substitutive competition." *Authors Guild*, 804 F.3d at 219. Here, the undisputed evidence shows that LLM training aims to extract unprotected elements from across a vast corpus of training data, rather than exploiting the particularized expression of ideas contained in any one work. *Supra* 17-18. Because this does not exploit the "precise form of expression," the commercial nature of the uses carries no weight. *See Consumers Union of U.S. v. Gen. Signal Corp.*, 724 F.2d 1044, 1049 (2d Cir. 1983).

B. The nature of the underlying works and the amount used carry little weight in light of the transformativeness of LLM training (Factors two and three).

Like Microsoft, Books Plaintiffs discuss the second and third factors only briefly. *See* Msft Books MSJ 24; BP MSJ 22-23, 33, 39. This is tacit agreement that these factors carry little weight here. Books Plaintiffs note that books are core copyrightable subject matter and that all (or at least large portions) of books are often acquired and used for LLM training. But this does not weigh against fair use where full copying of expressive works is "necessary to gain access to … unprotected functional elements." *Sony Comput. Ent., Inc. v. Connectix Corp.*, 203 F.3d 596, 603 (9th Cir. 2000). Undisputed evidence establishes that such copying is reasonably necessary to train an LLM, *see, e.g.*, Msft Books SUMF ¶¶ 7-9; OAI SUMF ¶¶ 9, 23; Defs' Books Counter-SUMF ¶¶ 471, 487-88, so the second and third factors weigh modestly in favor of fair use.

C. Training an LLM does not cause substantial market harm in any cognizable market (Factor four).

Books Plaintiffs' factor-four arguments confirm that the use of books to train an LLM is fair use as a matter of law. They fail to identify *any* harm they are suffering, let alone a legally cognizable harm in a traditional or likely to be developed market.

1. Undisputed material facts establish the absence of cognizable harm in any traditional book market.

Under factor four, not all markets and not all types of harm count. As *Texaco* explains, the only markets that matter are those that are "traditional, reasonable, or likely to be developed." 60 F.3d at 930. And the only cognizable harm ever recognized by the Supreme Court, Second Circuit, or any appellate court for that matter is substitution. *Authors Guild* defines this as "bring[ing] to the marketplace a competing

substitute for the original, or its derivative, so as to deprive the rights holder of *significant revenues* because of the likelihood that potential purchasers may opt to acquire the *copy in preference to the original*." 804 F.3d at 223 (emphasis added).

Books Plaintiffs' traditional markets are undisputed. They sell copies of books to readers through publishers. Msft Books SUMF ¶¶ 18, 21. And some generate revenue through audiobooks, distribution of books abroad, and for film and television adaptations. *Id.*; BP SUMF ¶¶ 6, 15, 46. LLM training does not cause cognizable harm in those markets for the simple reason that OpenAI's acquisition and use of books as training data is wholly unrelated to those markets. OpenAI is not and was never a would-be-reader or would-be-screenplay-adapter in the market for books. OpenAI's uses occur in an entirely "distinct market[]," *Oracle*, 593 U.S. at 37, that Books Plaintiffs could not even have fathomed when they authored their works.

Nor do the GPT models that result from the training process flood the market with copies (or derivatives) of Books Plaintiffs' books that consumers buy in preference to the originals. The record is stark on this point: The release of the GPT models has had no effect on Books Plaintiffs' sales or licensing opportunities. Individual Books Plaintiffs testified repeatedly that they could identify no lost sales. Msft Books SUMF ¶ 22; OAI SUMF ¶¶ 106-19. And expert analysis of 137 asserted works showed no discernible change in sales trends after the GPT models' release. Msft Books SUMF ¶¶ 23-24.

This lack of market harm is precisely what one would expect from a use as transformative as LLM training. Courts have repeatedly observed that such uses "do not serve as substitutes for the original work." *Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87, 99 (2d Cir. 2014). That is the simple reality of what is at issue in this case: A new

and different use yielding an innovative, general-purpose technology that does not hurt copyright owners.

2. Plaintiffs identify no cognizable market harm from any challenged use.

a. Plaintiffs' claims of harm from downloading or transfer of training data are empty.

Books Plaintiffs claim they suffer market harm from OpenAI's downloading from "pirate sites," *supra* 13, and any transfers of training data between OpenAI and Microsoft, *supra* 22-23. *See* BP MSJ 23-25 (acquisitions), 38-39 (distributions). The argument is simplistic. Returning to the theme of OpenAI as a dorm-room pirate engaging in "unauthorized, unpaid downloads," *id.* at 23, Plaintiffs argue that "[i]f anyone could lawfully obtain any book they wanted for free, there would be no reason … to pay for content," *id.* at 24 (cleaned up). So, they conclude, it must be that any use that a defendant does not pay for is substitutive.

The flaw in this perfect circle of an argument is the same one that infects all of Plaintiffs' arguments: They ignore that the new, different, non-traditional purpose of all of the challenged data acquisitions is training an LLM. When OpenAI downloaded books, it was not in the market for reading material. The same goes for Microsoft when it obtained temporary access to training data to evaluate the finished GPT-3 model. *Supra* 22-23. These uses were all quite outside of traditional book markets, and therefore did not substitute within those markets. *Authors Guild* reached a similar conclusion with respect to Google's copying of 20 million books in that case; because that use was transformative, the otherwise unauthorized creation of copies of those books caused no cognizable harm in traditional book markets. 804 F.3d at 223-24.

Contrast this with *Hachette*, the case Plaintiffs again rely upon. There, the defendants engaged in the same type of copying as in *Authors Guild* (scanning books), but for the non-transformative purpose of distributing those books. 115 F.4th at 195. The only way to reconcile these two cases, once again, is that *Authors Guild* involved a transformative purpose, and "the same copying may be fair when used for one purpose but not another." *Warhol*, 598 U.S. at 533; *see Kadrey*, 788 F. Supp. 3d at 1057 (rejecting the same theory of downloading market harm on this basis).

That same principle refutes Books Plaintiffs' argument that if OpenAI's downloading were "unrestricted and widespread," "conduct like OpenAI's unpaid downloads from pirate sites would" harm book markets. BP MSJ 24 (quotation marks omitted). They point to evidence that "piracy annually costs authors hundreds of millions of dollars in lost sales." *Id.* But no one is claiming that "piracy" in general is fair use. This case is about whether acquiring and using books to train LLMs is fair use. Again, same for Microsoft's temporary access to GPT-3 training data for purposes of evaluating and testing finished models. Permitting that sort of back-end engineering evaluation would hardly authorize widespread piracy for any purpose.

Last, Books Plaintiffs argue that downloading from LibGen "normaliz[es] piracy" and "encourag[es] others to use shadow libraries." *Id.* Plaintiffs make much of the fact that Microsoft's CEO testified, when asked about dorm-room-style content piracy, that Microsoft opposes such conduct. *See* Simpson Ex. 306, 175:13-176:6. What business leader would condone it? But again, OpenAI's downloading was for an entirely different purpose. The risk that OpenAI's past use of shadow library content might set a bad example is not a cognizable harm under factor four. *Cf. Authors Guild*, 804 F.3d at 223-24. And in any event, the law affords Books Plaintiffs every ability to directly pursue

shadow libraries or basic book piracy, as some already have. *See, e.g.*, Simpson Ex. 313, 50:25-52:19 (filed lawsuits and sends DMCA takedown notices); Simpson Ex. 314, 111:12-112:14 (filed lawsuit); Simpson Ex. 315, 68:6-69:16 (takedown notices); Simpson Ex. 316, 144:3-13 (same); Simpson Ex. 317, 161:6-11 (same); Simpson Ex. 318 at 3 (Authors Guild assisting with criminal copyright enforcement).

b. Plaintiffs' market dilution theory is legally untenable and factually speculative.

Books Plaintiffs are otherwise all in on the "market dilution" theory first hypothesized in *Kadrey*. Under that theory, "outputs from OpenAI and Microsoft's GPT-powered products 'can … compete for sales with the books in the training data,'" thus "reduc[ing] the incentive for authors to create." BP MSJ 34 (quotation marks omitted). As Microsoft explained in its motion (at 30-31), the market-dilution theory is invalid as a matter of law. And even if it were viable, Books Plaintiffs' lack of evidentiary compels the same result as in *Kadrey*.

Kadrey*'s market dilution theory is legally erroneous.* Books Plaintiffs do not explain precisely how their version of market dilution works. But under *Kadrey*'s conception, the argument goes like this. Books are used to train LLMs. These tokens supply inputs to the LLM's model weights, which aid the LLM in generating text. These LLMs can then be built into LLM-based products, which people can use to write. Some people might use these new tools to write new, non-infringing books. And some readers might *value* these new creative works, deriving entertainment or enrichment from them. Theoretically, it may even be possible that some of these people would buy one or more of these new books, which would deplete their book-buying budget such that they would not also buy a copy of, say, *The Lincoln Lawyer* by Michael Connelly.

According to the market-dilution theory, if the LLM developer used *The Lincoln Lawyer*'s tokens among the many trillions used to train the LLM that powered the tool that the new author used to write their book, then Michael Connelly can claim that the use of his book in training caused an indirect, non-infringing market substitute for *The Lincoln Lawyer*—and this weighs against fair use. If, it turns out, *The Lincoln Lawyer* is *not* in the training set, but the new author still writes and sells the same exact book using the otherwise identical LLM-based tool, Connelly is out of luck.

Neither *Kadrey* nor Plaintiffs cite any other case endorsing this theory. Indeed, *Kadrey* admits that "it's never made a difference in a case before." 788 F. Supp. 3d at 1055. *Kadrey* buttresses its theory by citing *Oracle* for the proposition that "[f]air use is meant to be a flexible doctrine that takes account of 'significant changes in technology,'" asserting that "[c]ourts can't stick their heads in the sand to an obvious way that a new technology might severely harm the incentive to create." *Id.* (quoting 593 U.S. at 19). Plaintiffs seize on this last line, suggesting that any use that "reduce[s] the incentive for authors to create" must be a cognizable harm, BP MSJ 34—while continuing to ignore the boon to original creation LLMs yield on the other side of the scale.

In any case, far from supporting *Kadrey*'s new theory, *Oracle* eviscerates the argument that just because a "new technology" may indirectly reduce the "incentive to create," that harm becomes cognizable under factor four. *Oracle* makes clear that not all "source[s] of … loss" count, and that some types of uses could "kil[l] demand for the original" and "[y]et …, even if directly translated into foregone dollars, … not [be] 'cognizable under the Copyright Act.'" 593 U.S. at 35 (citation omitted); *see also On Davis v. The Gap, Inc.*, 246 F.3d 152, 175 (2d Cir. 2001) ("[W]hen secondary uses harm the market for, or value of, the original, courts must examine the source of the harm.");

NXIVM Corp. v. Ross Inst., 364 F.3d 471, 482 (2d Cir. 2004) ("That the fair use, being transformative, might well harm, or even destroy, the market for the original is of no concern … so long as the harm stems from" something other than providing substitute copies).

For two centuries, the only cognizable source of loss under factor four has been from "supersed[ing] the use of the original work." *Folsom v. Marsh*, 9 F. Cas. 342, 345, 348 (C.C.D. Mass. 1841). "The fourth factor is aimed at the copier who attempts to *usurp* the demand for the original work" and is "not aimed at recompensing damages which may flow indirectly from copying." *Consumers Union*, 724 F.2d at 1050 (emphasis added). "[A] use which simply enables the copier to enter the market for works *of the same type* as the copied work" does not cause cognizable market harm. *Sega*, 977 F.2d at 1523 (emphasis added); *Arica Inst., Inc. v. Palmer*, 970 F.2d 1067, 1078 (2d Cir. 1992) ("Defendant … is perfectly entitled to create a competing work."); 4 Patry on Copyright § 10:150 ("[H]arm caused solely by the fact that the … parties' works compete in the same general market does not fall within the scope of Section 107(4)."). Any "harm" that "results from competition with an independently created work rather than from exploitation of plaintiff's own copyrighted material" is not cognizable. *Triangle Publ'ns, Inc. v. Knight-Ridder Newspapers, Inc.*, 626 F.2d 1171, 1178 (5th Cir. 1980) (citation omitted).

The law's disfavor only of uses that supersede or usurp the original—as opposed to compete with, criticize, comment on, or otherwise indirectly harm it—is rooted in the way copyright aims to incentivize creation. Copyright does not guarantee *a return* for authors; it guarantees only *exclusivity* in reproducing, distributing, and displaying one's own works to sell in traditional markets. Uses that usurp or supersede the original

interfere with those exclusive rights in those traditional markets, and therefore cause copyright-cognizable harm. Uses that merely compete with the original do not interfere with exclusive rights—the only form of incentive copyright law affords—and therefore are not cognizable. *Supra* 29-30.

At bottom, Plaintiffs' market dilution theory is based on the fear that a transformative use of their works may allow new authors to use a new technology to create new original works that people like. But copyright law venerates that sort of creative competition and rejects the "attempt to monopolize the market by making it impossible for others to compete" as "counter to the statutory purpose of promoting creative expression." *Sega*, 977 F.2d at 1523-24. Courts may not rewrite that statutory purpose by inventing new types of cognizable harms.

Plaintiffs fail to establish market dilution. Even if market dilution were a cognizable theory, Books Plaintiffs show no "meaningful or significant effect" from the use of their works to train the GPT models. *Authors Guild*, 804 F.3d at 224.

As *Kadrey* explained, "where a defendant introduces evidence of a lack of market harm, 'and the plaintiff fails to introduce empirical evidence countering such a showing, the fourth factor should be weighed in the defendant's favor.'" 788 F. Supp. 3d at 1056 (citing Patry on Fair Use § 6:13). Microsoft has done that here, providing undisputed evidence that Books Plaintiffs' sales have been unaffected by the introduction of the GPT models. *Supra* 6. It therefore falls to Books Plaintiffs to answer the empirical questions *Kadrey* posed in order to establish market dilution: "[W]hat are these AI-generated books?" "Do they compete with Sarah Silverman's memoir? With plaintiff Matthew Klam's book of short stories?" "[W]hat impact does this competition actually have"? "[D]oes it depend on the book"? "[H]ow does the threat … compare" to an LLM not

trained on a book? 788 F. Supp. 3d at 1056. All of these showings are essential to proving "the effect of defendants' use of [plaintiffs'] expression on plaintiffs' market." *NXIVM*, 364 F.3d at 482 (emphasis added).

Even with *Kadrey*'s guide, Books Plaintiffs come up empty. Their single-paragraph set of generalities about self-published works on Amazon does not account for how Amazon defines the use of AI in those works. For example, Plaintiffs claim that [redacted] BP MSJ 36. But the cited evidence reflects [redacted] Defs' Books Counter-SUMF ¶ 849. *Books Plaintiffs* themselves have used generative AI in this fashion. *See, e.g.*, Simpson Ex. 313, 91:15-22 ("editing [her] books," "designing the cover art for the books," "doing translations for the books"); Simpson Ex. 320, 81:22-82:11 ("streamline something or make it shorter"). Books Plaintiffs' assertion that books "can be generated in less than ten minutes and for less than a dollar," BP MSJ 36, assumes a book wholly created by AI, Defs' Books Counter-SUMF ¶¶ 845-46, which bears no relationship to the Amazon data, *id.* ¶¶ 849-54.

Books Plaintiffs also have no evidence for the claim that "AI books are competing at the highest levels of sales." BP MSJ 36; Defs' Books Counter-SUMF ¶¶ 864-68. Books Plaintiffs withdrew the lone study they cited in support after Defendants pointed to evidence that Books Plaintiffs' own lawyers in this case had funded the research. ECF 2087 at 2; ECF 2090; Defs' Books Counter-SUMF ¶¶ 858-71. And Microsoft's economic expert explained at length why mere evidence of an increase in supply of books with AI-

generated text, translations, or images does not mean that such books would meaningfully compete with Books Plaintiffs. Tucker ¶¶ 61-76. This Court should demand more than speculative harm under a legally dubious theory. *Supra* 31; *iParadigms*, 562 F.3d at 644-45 (rejecting plaintiffs' theories of "potential market effects" as "unfounded speculation").

c. There is no legally cognizable market for training data.

Books Plaintiffs' last market-harm argument is that "OpenAI's unlicensed use deprived Plaintiffs of remuneration from an existing licensing market." BP MSJ 36 (caps omitted). Defendants have already explained at length why this argument cannot work as a matter of law or fact. Msft Books MSJ 27-30. As *Kadrey* explained, whether a market for training data "exists or is likely to develop is irrelevant, because this market is not one that plaintiffs are legally entitled to monopolize." 788 F. Supp. 3d at 1052. *Bartz* held the same. 787 F. Supp. 3d at 1032. The Second Circuit rejected the same theory in *HathiTrust*, holding that plaintiffs cannot claim a "lost sale" from the very transformative use they are challenging. 755 F.3d at 99-100. It has also held that a copyright owner cannot "preempt exploitation of transformative markets … by actually developing or licensing others to develop those markets." *Castle Rock Ent., Inc. v. Carol Publ'g Grp., Inc.*, 150 F.3d 132, 145 n.11 (2d Cir. 1998). Because Books Plaintiffs make no attempt to meet this fatal legal challenge, we need not belabor the point.

On the (legally immaterial) facts, Books Plaintiffs point to numerous licensing deals for training data, while again observing that Defendants recognized the value of training on books generally. BP MSJ 37. They claim this shows that there is now a "robust," "growing," and "sustainable" market. *Id.* at 37-38. But the fact that LLM

developers have tried to figure out a way to license some training data does not mean that these efforts have been successful, *see* Usdan ¶¶ 23-28—and certainly not that there is a "reasonably efficient, reasonably priced, [and] convenient [to] access" licensing market for sufficient LLM training data, *Cambridge Univ. Press v. Patton*, 769 F.3d 1232, 1279 (11th Cir. 2014).

That is where the licensing-market argument falls apart. It is undisputed that training requires trillions upon trillions of tokens, and that any one author could supply only a tiny fraction. Msft Books SUMF ¶¶ 5-6, 10-11, 28(a). So too that copyright protection subsists in every work of authorship automatically, § 102(a), and that these millions of authors are dispersed, Msft Books SUMF ¶ 28(e). And once they are found, negotiations are costly and holdup inevitable, in part because there is no viable way of valuing any single piece of training data. *Id.* ¶ 28(c), (f)-(g). Case in point: [redacted] which Books Plaintiffs highlight. BP MSJ 37. That experiment was "a failure." Usdan ¶ 24. "Negotiations took a very long time." *Id.* Even then the agreement yielded only [redacted] *Id.* ¶ 22. And these books wound up contributing very little to actual model development. *Id.* ¶ 24(a). As a result of the lessons learned, Microsoft [redacted] *Id.* ¶ 25.

In short, even if a market for training data were relevant, undisputed evidence shows at a minimum that there are enormous—and very probably insurmountable—barriers to its development.

***

For reasons explained in Microsoft's motion and above, this Court should find that LLM training is fair use as a matter of law. The facts that are actually material

demonstrate that this use is immensely transformative and has caused no market impact. Books Plaintiffs' motion should be denied and Microsoft's motion granted.

II. Plaintiffs Fail To Establish Microsoft's Copying Or Distribution As To The Vast Majority Of Direct Infringement Claims.

The Court should also deny Books Plaintiffs' motion as to direct infringement claims against Microsoft for the vast majority of alleged copies or distributions. As noted above (at 13), a direct infringer is one who engages in the "volitional conduct that cause[d] the copy to be made." *Cablevision*, 536 F.3d at 131. It is only the actor who "actually 'makes' a copy" (or distribution) and thus "commits a copyright-infringing act" that can be held liable for direct infringement. *Id.* at 131, 133. Books Plaintiffs move for summary judgment on all copies and distributions in this case, while ambiguously claiming that "OpenAI and Microsoft both copied … in at least three separate ways." BP MSJ 18. But they point to no evidence that Microsoft made copies or distributions as to all but the handful of transfers described above (at 22-23), which contained few if any of Books Plaintiffs' works. Apart from these, summary judgment should be denied, and in the case of downloads from shadow libraries Microsoft's own motion should be granted.

Books Plaintiffs adduce no evidence that Microsoft engaged in volitional conduct as to "downloading … from LibGen and Books3." BP MSJ 18-19. Insofar as Books Plaintiffs predicate direct infringement claims on the fact that Microsoft provided the supercomputing infrastructure and cloud storage that OpenAI used for data acquisition and storage, that argument is foreclosed by *Cablevision*. There is no evidence that Microsoft "actually presse[d] the button" or was otherwise directly involved in OpenAI's downloads from pirate sites, some of which took place well before Microsoft entered the

picture. 536 F.3d at 131; Msft Books SUMF ¶¶ 30-32; Defs' Books Counter-SUMF ¶¶ 160, 180-83, 633.

There is also no evidence that Microsoft was the volitional actor with respect to "OpenAI['s] … additional reproductions of Plaintiffs' books in training its LLMs." BP MSJ 19. OpenAI performed all challenged conduct in training the GPT models at issue in this case. *Supra* 4-5, 13; *infra* 37-40. Summary judgment against Microsoft is therefore inappropriate for any such conduct. *See* OAI SUMF ¶¶ 31-32, 35-38, 141, 343; Defs' Books Counter-SUMF ¶¶ 438, 517-31, 577, 659. And finally, with limited and specific exceptions discussed above (at 22), Microsoft was not the volitional actor as to many alleged distributions of "copies of Plaintiffs' works," BP MSJ 19. *See* Defs' Books Counter-SUMF ¶¶ 572, 607-08. Summary judgment should be denied as to these as well.

III. Books Plaintiffs Are Not Entitled To Summary Judgment That Microsoft Is Vicariously Liable For OpenAI's Direct Infringement.

Finally, Books Plaintiffs seek a sweeping ruling that "Microsoft is vicariously liable for OpenAI's infringement." BP MSJ 40 (caps omitted). They appear to be arguing vicarious liability as to every single act by OpenAI at issue. To support the extraordinary request to hold one independent company liable, as a matter of law, for all acts of another independent company over a several year period, Books Plaintiffs offer a total of two conclusory paragraphs. That perfunctory showing alone merits denying their motion. *See In re Demetriades*, 58 F.4th 37, 54 (2d Cir. 2023); *United States ex rel. Taylor v. GMI USA Corp.*, 774 F. Supp. 3d 645, 672 (S.D.N.Y. 2025).

Even if this Court is inclined to piece together disparate record cites and guess at their purported legal significance, the record does not support summary judgment. In

addition to a showing of direct infringement, *but see supra* 9-35, Books Plaintiffs bear the burden of showing that Microsoft "had a right and ability to supervise that coalesced with an obvious and direct financial interest in the exploitation of copyrighted materials" by OpenAI. *Softel, Inc. v. Dragon Med. & Sci. Commc'ns, Inc.*, 118 F.3d 955, 971 (2d Cir. 1997) (citation modified). Determinations as to both are often "fact-specific," and they are here. *See Adobe Sys. Inc. v. Canus Prods., Inc.*, 173 F. Supp. 2d 1044, 1053, 1055 (C.D. Cal. 2001) (denying plaintiff's motions on vicarious liability). The evidence cannot establish either element.

A. Plaintiffs fail to show that Microsoft had the legal right and practical ability to supervise OpenAI's data acquisition and training activities.

1. As an outgrowth of "the doctrine of respondeat superior," vicarious liability requires a showing that a defendant has the "right and ability to supervise" a direct infringer. *Shapiro, Bernstein & Co. v. H.L. Green Co.*, 316 F.2d 304, 307 (2d Cir. 1963). A plaintiff must demonstrate both a "legal right to stop or limit the directly infringing conduct, as well as the practical ability to do so." *Perfect 10, Inc. v. Amazon.com, Inc.*, 508 F.3d 1146, 1173 (9th Cir. 2007). A "formal relationship between parties is not the driving force behind liability." *Banff Ltd. v. Ltd., Inc.*, 869 F. Supp. 1103, 1109 (S.D.N.Y. 1994). What matters is that the defendant "is in a position to control the personnel and activities responsible for the direct infringement." *Id.*

Witness after witness for both OpenAI and Microsoft confirmed that training the challenged GPT models was OpenAI's exclusive domain, and Microsoft's role was as a business and developmental partner that supplied capital and infrastructure. A core "ten[e]t of the OpenAI/Microsoft relationship … [was] that OpenAI … steer[s] the technology." Simpson Ex. 312, 169:24-170:17. Microsoft's role was limited to providing

"an investment in a third-party company that [Microsoft] [did not] control [or] operate," Simpson Ex. 321, 20:19-24, and "a computing platform for [OpenAI] to use," Simpson Ex. 322, 32:18-33:6. "[T]he decision on what data was used [to train LLMs] was 100 percent OpenAI's decision," Simpson Ex. 323, 167:9-10, and Microsoft "didn't have control over what training data went into the models themselves," Simpson Ex. 310, 248:14-20, nor "any involvement … in the pretraining decision or pretraining execution," "design decisions or assisting with engineering," or "the design, decision-making, or execution of relevant runs," Simpson Ex. 312, 162:23-163:14, 165:2-166:2.

2. Books Plaintiffs nevertheless ask for summary judgment based on (a) provisions of the joint development and collaboration agreements, or JDCAs, governing OpenAI and Microsoft's partnership; and (b) Microsoft's purported "control over the Azure servers" that OpenAI used for its training activities. BP MSJ 40.

Books Plaintiffs misread the JDCAs. To begin with, the agreements expressly disclaim "agency among the parties," and therefore must be interpreted in that light. Simpson Ex. 324 at -366; Simpson Ex. 325 at -608. They also say that "OpenAI may train all Models using data selected *solely* by … OpenAI." Simpson Ex. 325 at -590 (emphasis added); *see also id.* at -644 ("OpenAI may use any datasets to train Models"). Other JDCA provisions confirm OpenAI's exclusive authority over LLM training. *See* Simpson Ex. 324 at -354 [redacted] (emphasis added)); *id.* at -352 [redacted] (emphasis added)); *id.* at -348-49 [redacted] (emphasis added)).

Plaintiffs point to the JDCAs' creation of a "Governing Board," BP MSJ 8, misleadingly truncating a provision which actually says that the Governing Board is responsible for "monitoring and controlling the parties' day-to-day operations *under this Agreement*," Simpson Ex. 326 at -4102 (emphasis added). The agreements dictate no "day-to-day operations" related to data acquisition and training, a function for which OpenAI was responsible. They are about other aspects of the parties' partnership, like "training infrastructure" and "IP collaborations." Simpson Ex. 327, 393:12-19; Simpson Ex. 329, 116:17-24. Undisputed testimony confirms that the Governing Board operated accordingly, never involving itself in OpenAI's data acquisition or training (let alone day-to-day). *See* Simpson Ex. 328, 68:12-21; Simpson Ex. 329, 116:11-24. In any case, the Governing Board was limited to providing non-binding "advice and recommendations." Defs' Books Counter-SUMF ¶ 645.

Last, Books Plaintiffs highlight a clause from post-January 2023 JDCAs giving Microsoft a "(never-exercised) right to require OpenAI to 'promptly retrain [a] Model using solely data approved by Microsoft' if Microsoft had 'concerns regarding the use of any particular data to train a Model.'" BP MSJ 8; Defs' Books Counter-SUMF ¶ 690. This is a non-starter because that clause was not yet in place when OpenAI trained models using data from shadow libraries, which was limited to GPT-3 and GPT-3.5. *See* OAI SUMF ¶ 11; BP SUMF ¶¶ 519-21, 526. The retraining clause was merely a remedial measure anyway. It gave Microsoft no right to monitor OpenAI's conduct on an ongoing basis. *See also* Simpson Ex. 323, 125:17-126:19.

3. Books Plaintiffs get no further with the suggestion that Microsoft "had control over the Azure servers on which OpenAI downloaded its data, stored its data, trained its models, and deployed its models." BP MSJ 40. Microsoft's provision of cloud

infrastructure does not give it day-to-day control over OpenAI's activities, any more than over the countless other organizations that use Microsoft's Azure infrastructure. *See* Defs' Books Counter-SUMF ¶ 679. Microsoft's technical policies did not allow it to access customer data stored on Azure, including the contents of OpenAI's training data or processes on Azure systems. *See* Simpson Ex. 335, 65:16-66:19. Microsoft's monitoring of OpenAI's use of Azure [redacted] Simpson Ex. 332, 46:1-9.

Cases have repeatedly rejected liability based on provision of servers a defendant had no right or ability to actively monitor or limit. *In re Google Generative AI Copyright Litig.*, 809 F. Supp. 3d 903, 917 (N.D. Cal. 2025) (lack of "technological ability to search for or recognize Plaintiffs' works during [defendant's] selection or use of the[] massive training datasets"); *see Luvdarts, LLC v. AT&T Mobility, LLC*, 710 F.3d 1068, 1071 (9th Cir. 2013) (defendant had "no way of supervising the use of their networks for copyright infringement"); *VHT, Inc. v. Zillow Grp., Inc.*, 918 F.3d 723, 746 (9th Cir. 2019) ("insufficient evidence that [defendant] had the technical ability to screen out or identify infringing … photos"). And this case is nothing like the one Books Plaintiffs cite, *EMI Christian Music Group, Inc. v. MP3tunes, LLC*, in which a CEO was held liable for his own company's infringement based on granular control and visibility into that conduct. 844 F.3d 79, 99-100 (2d Cir. 2016).

B. Plaintiffs fail to show that Microsoft derived a direct financial benefit from OpenAI's alleged infringement of their works.

Though Books Plaintiffs identify three categories of Microsoft revenue related to OpenAI, they articulate no "causal relationship between *the infringing activity* and

[each] financial benefit." *Cengage Learning Inc. v. Google LLC*, 786 F. Supp. 3d 611, 621 (S.D.N.Y. 2025) (citation omitted) (emphasis added).

First, Books Plaintiffs point to revenue that Microsoft received from OpenAI's use of Azure for LLM training. BP MSJ 40. Plaintiffs offer no evidence that OpenAI's use of allegedly infringing datasets to train LLMs had a differential impact on the Azure revenue it received. Defs' Books Counter-SUMF ¶¶ 644, 646, 703; *see Sony Music Ent. v. Cox Commc'ns, Inc.*, 93 F.4th 222, 231 (4th Cir. 2024) (receiving "flat periodic payments for service from infringing and noninfringing use[s] alike ordinarily would not constitute receiving a financial benefit directly attributable to the infringing activity" (citation modified)), *rev'd and remanded on other grounds*, 607 U.S. 583 (2026). Nor have Plaintiffs established that Microsoft "received a direct financial benefit from the infringement … of [*Plaintiffs'*] copyrighted material, rather than general infringement." *Perfect 10, Inc. v. Giganews, Inc.*, 847 F.3d 657, 673 (9th Cir. 2017).

Next, Plaintiffs invoke the share of OpenAI's overall revenue that Microsoft received. BP MSJ 40. That share was a negotiated, contractually fixed return on Microsoft's capital commitment, not compensation keyed to any dataset or training data. Defs' Books Counter-SUMF ¶¶ 125, 634, 701; Simpson Ex. 325 at -596; *Ellison v. Robertson*, 357 F.3d 1072, 1079 (9th Cir. 2004). Nor did Plaintiffs even attempt to prove that the Asserted Works, as opposed to "books as a class," Zhao Ex. A ¶ 72, had any discernible value with respect to the challenged models' performance, *see* Usdan ¶ 10. It is thus entirely unclear based on Books Plaintiffs' cited evidence whether there is any direct relationship between OpenAI's use of their works for training and Microsoft's share of OpenAI's revenue. *See Michael Grecco Prods., Inc. v. Valuewalk, LLC*, 345 F. Supp. 3d 482, 504 (S.D.N.Y. 2018) (denying summary judgment where "the

parties dispute whether the allegedly infringing image itself, rather than the text of the page, was used to attract readers in an effort increase revenue").

Even more attenuated is Plaintiffs' final theory that Microsoft enjoyed incrementally higher Azure and Copilot subscribers from uses of GPT models. BP MSJ 40. Books Plaintiffs point to no evidence about Microsoft's standalone Azure and Copilot product revenue at all, let alone proof that more subscribers were drawn to Azure and Copilot because the GPT models were trained on Plaintiffs' works amongst the trillions of other tokens used to train those models. Defs' Books Counter-SUMF ¶ 125.

CONCLUSION

For the foregoing reasons, Books Plaintiffs' motion should be denied.

Dated: October 23, 2026

Respectfully submitted,

Annette L. Hurst

ORRICK, HERRINGTON & SUTCLIFFE LLP
Annette L. Hurst (admitted pro hac vice)
Daniel Justice (admitted pro hac vice)
Robert Scott 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 Opposition complies with the Court's order of August 7, 2026 (ECF 1664), as it contains 12,461 words, according to the word processing program used to prepare it.

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