

**UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK**

IN RE OPENAI, INC., COPYRIGHT
INFRINGEMENT LITIGATION

This document relates to:

No. 1:23-cv-08292 (SHS) (OTW)

No. 1:23-cv-10211 (SHS) (OTW)

No. 1:25-cv-03482 (SHS) (OTW)

Case No. 25-md-3143-SHS-OTW

Hon. Sidney H. Stein

FILED UNDER SEAL

**CLASS PLAINTIFFS' MEMORANDUM OF LAW IN SUPPORT OF MOTION FOR
PARTIAL SUMMARY JUDGMENT**

TABLE OF CONTENTS

INTRODUCTION.....	1
FACTUAL BACKGROUND	4
I. Plaintiffs and Their Works.	4
II. OpenAI Has Become a \$1 Trillion Business Built On Its GPT-Series of LLMs.....	4
III. OpenAI Downloaded Millions of Books Without Payment.	5
A. OpenAI Torrented More Than 4 Million Books From The Pirate Site LibGen.	5
B. OpenAI Downloaded At Least 21,000 Books from the Pirated Books3 Dataset.	7
C. OpenAI Downloaded Large Quantities of Books from Other Sources.	7
D. OpenAI And Microsoft Worked Together to Download Large Numbers of Books Via Crawling.....	7
IV. The OpenAI-Microsoft Partnership.....	8
A. The Joint Development and Collaboration Agreements.....	8
B. OpenAI and Microsoft Used Books as Currency in Their Commercial Relationship and Distributed Books to Each Other.	9
C. OpenAI Distributed LibGen Books To Third Party Contractors.....	9
V. OpenAI and Microsoft Exploited Plaintiffs’ Works to Train the At-Issue LLMs to Displace Human Authors.	10
A. The LLM Training Process and the In-Suit LLMs.	10
B. OpenAI Copied Plaintiffs’ Books to Train In-Suit LLMs.....	11
C. OpenAI and Microsoft Valued and Licensed Books for Their Expressive Qualities. 12	
D. The In-Suit LLMs Reproduce Plaintiffs’ Copyrighted Material to End Users.....	13
E. OpenAI Designed Its LLMs to Replace Human Creations.	13
F. ██████████ Shows AI-Generated Works Are Rapidly Entering the Market and Gaining Sales.....	15
LEGAL STANDARD.....	16
I. Summary Judgment.....	16
II. Copyright Law	17
A. Direct Infringement.....	17
B. Vicarious Infringement	17
C. Fair Use.....	18
ARGUMENT.....	18
I. The Elements of a Copyright Infringement Claim Are Satisfied As a Matter of Law. 18	
A. Plaintiffs Own Valid Copyrights in Their Works.	18
B. OpenAI and Microsoft Copied Plaintiffs’ Works When Initially Downloading the Works, When Training Their LLMs, and When Distributing the Works To One Another. 18	

II. OpenAI’s Unauthorized Downloading of Plaintiffs’ Works from LibGen and Books3 Was Not Fair Use as a Matter of Law.	19
A. Factor One: OpenAI’s Downloading from Pirate Sites Was Nontransformative, Lacks Justification, and Is Blatantly Commercial.	19
B. Factor Two: Plaintiffs’ Copyrighted Books are at the Core of Copyright Protection.	22
C. Factor Three: OpenAI Pirated Plaintiffs’ Books in Their Entirety.	23
D. Factor Four: Piracy Harms the Market for Plaintiffs’ Copyrighted Books	23
III. Defendants’ Copying of Plaintiffs’ Works to Train LLMs Was Not Fair Use.....	25
A. Factor One: OpenAI’s Copying of Plaintiffs’ Books To Develop LLMs Was Non-Transformative, Lacks Any Independent Justification, and Was Commercial.	25
i) OpenAI Copied Plaintiffs’ Works To Develop A Product Designed To Generate Competing Output.	26
ii) OpenAI’s Copying of Plaintiffs’ Works Was Not “Reasonably Necessary” To Achieve Its Purportedly Transformative Purpose.	29
iii) OpenAI’s Copying Was Commercial.	32
B. Factors Two and Three Favor Plaintiffs: Plaintiffs’ Copyrighted Books are at the Core of Copyright Protection and OpenAI Copied Them In Their Entirety.	33
C. Factor Four: Defendants’ Use of Plaintiffs’ Works Threatens the Market For Plaintiffs’ Works and the Deprives Them of Licensing Revenues.	33
i) Defendants’ Use of Books in AI Training Threatens Wholesale Market Displacement of Human-Authored Books.	34
ii) OpenAI’s Unlicensed Use Deprived Plaintiffs of Remuneration from an Existing Licensing Market.	36
IV. Defendants’ Distribution of Plaintiffs’ Works to Each Other and Third Parties Was Not Fair Use.	38
A. Factor One: Defendants’ Transfers Were Substitutive, Not Transformative.	38
B. Factors Two & Three Favor Plaintiffs.	39
C. Factor Four: Unlicensed Distribution of Plaintiffs’ Works Threatens Complete Destruction of Plaintiffs’ Traditional Markets.	39
V. Microsoft Is Vicariously Liable for OpenAI’s Infringement.	40
CONCLUSION.....	40

TABLE OF AUTHORITIES

	Page(s)
Cases	
<i>A&M Recs., Inc. v. Napster, Inc.</i> , 239 F.3d 1004 (9th Cir. 2001)	2, 19, 20, 23
<i>Am. Geophysical Union v. Texaco Inc.</i> , 60 F.3d 913 (2d Cir. 1994).....	<i>passim</i>
<i>Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith</i> , 598 U.S. 508 (2023).....	<i>passim</i>
<i>Arista Records, LLC v. Doe 3</i> , 604 F.3d 110 (2d Cir. 2010).....	17, 24
<i>Arista Recs., Inc. v. Mp3Board, Inc.</i> , 2002 WL 1997918 (S.D.N.Y. Aug. 29, 2002) (Stein, J.)	20, 39
<i>Arista Recs. LLC v. Usenet.com, Inc.</i> , 633 F. Supp. 2d 124 (S.D.N.Y. 2009).....	17, 19, 24
<i>Authors Guild, Inc. v. HathiTrust</i> , 755 F.3d 87 (2d Cir. 2014).....	27, 29
<i>Bartz v. Anthropic PBC</i> , 787 F. Supp. 3d 1007 (N.D. Cal. 2025)	<i>passim</i>
<i>Basic Books, Inc. v. Kinko's Graphics Corp.</i> , 758 F. Supp. 1522 (S.D.N.Y. 1991).....	21, 39
<i>BMG Music v. Gonzalez</i> , 430 F.3d 888 (7th Cir. 2005)	19, 20, 24
<i>Byrne v. Brit. Broad. Corp.</i> , 132 F. Supp. 2d 229 (S.D.N.Y. 2001) (Stein, J.)	23, 25, 31
<i>Campbell v. Acuff-Rose Music Inc.</i> , 510 U.S. 569 (1994).....	20, 22, 23
<i>Elsevier Inc. v. Sci-Hub</i> , No. 15-cv-4282, 2017 WL 3868800 (S.D.N.Y. June 21, 2017)	5
<i>EMI Christian Music Grp., Inc. v. MP3tunes, LLC</i> , 844 F.3d 79 (2d Cir. 2016).....	17, 40

<i>Fox News Network, LLC v. Tveyes, Inc.</i> , 883 F.3d 169 (2d Cir. 2018).....	19, 20, 21
<i>Google Books</i> , 804 F.3d 202	27, 34, 35
<i>Hachette Book Grp., Inc. v. Internet Archive</i> , 115 F.4th 163 (2d Cir. 2024)	<i>passim</i>
<i>Hachette Book Grp., Inc. v. Internet Archive</i> , 664 F. Supp. 3d 370 (S.D.N.Y. 2023), <i>aff'd</i> , 115 F.4th 163 (2d Cir. 2024).....	16, 17, 24, 25
<i>Harper & Row Publishers, Inc. v. Nation Enters.</i> , 471 U.S. 539, 568 (1985).....	24, 25
<i>Harris v. Simon & Schuster, Inc.</i> , 646 F. Supp. 2d 622 (S.D.N.Y. 2009).....	18
<i>In re Frontier Communications Corporation</i> , 658 B.R. 277 (Bankr. S.D.N.Y. 2024).....	17, 40
<i>Jorgensen v. Epic/Sony Records</i> , 351 F.3d 46 (2d Cir. 2003).....	17
<i>Kadrey v. Meta Platforms</i> , 788 F. Supp. 3d 1026, 1035 (N.D. Cal. 2025)	<i>passim</i>
<i>McGucken v. Pub Ocean Ltd. et al.</i> , 42 F.4th 1149 (9th Cir. 2022)	30
<i>O'Neil v. Ratajkowski</i> , 563 F. Supp. 3d 112 (S.D.N.Y. 2021).....	18
<i>Romanova v. Amilus Inc.</i> , 138 F.4th 104 (2d Cir. 2025)	29
<i>Thomson Reuters Enter. Ctr. GMBH v. Ross Intel. Inc.</i> , 765 F. Supp. 3d 382 (D. Del. 2025).....	27, 29
<i>UMG Recordings, Inc. v. MP3.Com, Inc.</i> , 92 F. Supp. 2d 349 (S.D.N.Y. 2000).....	37
<i>Warner Bros. Ent. Inc. v. RDR Books</i> , 575 F. Supp. 2d 513 (S.D.N.Y. 2008).....	23
<i>Williams v. Crichton</i> , 84 F.3d 581 (2d Cir. 1996).....	17

Statutes

17 U.S.C.A. § 501(b)	17
17 U.S.C. § 106(3)	38
17 U.S.C. § 107	18, 21, 34
17 U.S.C. § 107(1)	19
17 U.S.C. § 107(4)	23

Rules

Fed. R. Civ. P. 56(a)	16
-----------------------------	----

Constitutional Provisions

U.S. Const. Art. I § 8, cl. 8.....	25
------------------------------------	----

INTRODUCTION

“Our work on AI and Creativity is going to increasingly lead to us creating systems that substitute for the labor of [] people . . . The better we do on GPT-X, the more worried genre fiction authors will become about us substituting for them on Amazon.” OpenAI Policy Director Jack Clark, May 2020 (Statement of Undisputed Facts (“SUF”) ¶749).

“We trained GPT-3 on pirated stuff! No sharing that!” Internal OpenAI Notes, August 2019 (SUF¶344).

“Q: You agree that piracy is wrong and illegal? A: I do.” Sam Altman, February 2026 (SUF¶370).

“Q: It’s illegal to download pirated material, right? A: Absolutely. That’s correct.” Satya Nadella, February 2026 (SUF¶382).

OpenAI’s GPT models pose an existential threat to those who write and publish books. AI-generated books of all types are already flooding the market. These books threaten to “substitute” for the creations of authors, as OpenAI’s then-head of policy Jack Clark bluntly admitted on the eve of GPT-3’s release in May 2020. SUF¶749. OpenAI’s models are capable of displacing human authors precisely *because* of the unauthorized exploitation of their works. These models mine and mimic the protected expression in books, enabling Defendants’ commercial products to generate competing works at the press of a button. OpenAI did not even buy the books it used. Instead, it began by torrenting millions of books from the notorious and illegal pirate library Library Genesis, also known as LibGen. Microsoft encouraged, facilitated, and profited from this mass infringement, and had the ability to control it along the way. It is no exaggeration to say that OpenAI built the foundations of its business on mass piracy.

Defendants do not dispute that they copied Plaintiffs’ works and millions of others. Instead, they argue that their conduct is excused as fair use. But fair use does not countenance such catastrophic threats to the incentive to create.

“[T]he problem of substitution [is] copyright’s *bête noire*.” *Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith*, 598 U.S. 508, 528 (2023). In this case, Defendants’ conduct

implicated the problem of substitution at each step. Each instance of copying—analyzed independently, as required by *Warhol*—usurps the market for Plaintiffs’ books and threatens the very foundations of copyright.

First, OpenAI downloaded millions of books—including Plaintiffs’ books—from LibGen, a site recognized as a Notorious Market for Counterfeiting and Piracy by the U.S. Trade Representative. SUF¶133. This scene has played out in many cases before. In the wake of the Napster era, courts consistently held that obtaining “for free” something one would “ordinarily have to buy” is not fair use. *A&M Recs., Inc. v. Napster, Inc.*, 239 F.3d 1004, 1015 (9th Cir. 2001). “Such piracy of otherwise available copies is inherently, irredeemably infringing[.]” *Bartz v. Anthropic PBC*, 787 F. Supp. 3d 1007, 1025 (N.D. Cal. 2025).

Second, OpenAI made further copies of Plaintiffs’ works and used them to train commercial products capable of flooding the market with books. In conscripting Plaintiffs’ works for its GPT models, OpenAI consumed the creative expression—and used it to encode the word choice, syntax, turns of phrase, characters, and other protected elements into its model—with the express purpose of substituting for Plaintiffs and their works.

Worse still, the GPT models built on Plaintiffs’ works threaten immense market harm. In just the few years since ChatGPT’s release, AI-generated content accounts for [REDACTED] and such content is diluting the market for books across the board. SUF¶851. OpenAI’s models generate market-competitive books in minutes and at a fraction of the cost. The models possess these capabilities because they were trained on and are able to mimic high-quality creative expression like Plaintiffs’ works. SUF¶708, 734-35, 745-50, 845-47. As Judge Chhabria recognized in *Kadrey v. Meta Platforms*, “by training generative AI models with copyrighted works, companies are creating something that often will dramatically undermine the

market for those works, and thus dramatically undermine the incentive for human beings to create things the old-fashioned way.” 788 F. Supp. 3d 1026, 1035 (N.D. Cal. 2025).

The licensing market for AI training confirms that Defendants can compensate rightsholders even as they develop technology valued in the trillions. As the Second Circuit has recognized, if Defendants “want to continue the precise copying” at issue here, they can “use the licensing schemes now existing” without imperiling the incentive copyright exists to protect. *Am. Geophysical Union v. Texaco Inc.*, 60 F.3d 913, 932 (2d Cir. 1994).

Third, OpenAI and Microsoft traded compilations of Plaintiffs’ books via “horse-trading.” SUF¶590. Defendants treated large datasets of copyrighted text as currency exchanged for commercial concessions and other data. This copying and distribution too is not fair use.

Fourth, Microsoft is vicariously liable for OpenAI’s infringements as a matter of law. Microsoft knew about OpenAI’s use of LibGen as early as April 2019 when Sam Altman and Dario Amodei presented an early version of GPT-3 to Bill Gates and disclosed the use of LibGen to Gates and Microsoft’s Chief Technology Officer Kevin Scott, among others. SUF¶182-191, 519. Microsoft nevertheless invested in OpenAI, negotiating the right and ability to supervise OpenAI’s infringing conduct all while using Microsoft hardware and services and powering Microsoft to new levels of profitability as a leader in AI.

* * *

Plaintiffs are some of the most prolific and well-known authors in the world. They collectively have written some of the most consequential books of the last half century. These creations exist because of the copyright protections enshrined in the Constitution and the Copyright Act. But this case is not just about Plaintiffs’ books. Defendants’ blatant copyright violations threatens the incentives to create, the livelihoods of authors, and the future of book publishing.

Defendants’ unauthorized use of Plaintiffs’ works—if left unchecked and uncompensated—endangers the market for books. For good reason, the doctrine of fair use does not protect Defendants’ wanton infringement.

FACTUAL BACKGROUND

I. Plaintiffs and Their Works.¹

Plaintiffs are bestselling and award-winning book authors² and The Authors Guild, which is the oldest and largest professional organization for published writers in the United States. Plaintiffs have spent years of their lives honing their craft, researching and writing their books, and seeking an audience for their life’s work. Although this summary judgment motion is about their works, Plaintiffs are proud to represent the interests of copyright holders in all types of books—authors and publishers alike—to vindicate the basic principles which Defendants’ conduct threatens. SUF¶2.

This motion addresses 194 of Plaintiffs’ works (“Asserted Works”). These Asserted Works span a wide range of genres, styles, subject matters, and sales volumes, including literary fiction, genre fiction, drama, history, and biography. Collectively these works have sold hundreds of millions of copies, won Pulitzer prizes, and been adapted into award-winning movies and television shows. The Asserted Works are listed in Appendix A to Plaintiffs’ Statement of Undisputed Material Facts.

II. OpenAI Has Become a \$1 Trillion Business Built On Its GPT-Series of LLMs.

Founded in December 2015, OpenAI began as a tax-exempt nonprofit. SUF¶85. Despite its nonprofit origin story, OpenAI and its founders had commercial ambitions from the outset. Co-

¹ Plaintiffs’ Statement of Undisputed Material Facts, filed alongside with this motion, contains a full recitation of the relevant material facts.

² A number of the named Plaintiffs are entities through which authors conduct professional activities, called “loan-out companies.” SUF¶1.

founder and President Greg Brockman wrote in 2017 that he was “deeply motivated by the gazillions” OpenAI could make as a for-profit enterprise. *SUF*¶89. OpenAI changed to a for-profit company in 2019. *SUF*¶87. Today, OpenAI is preparing for an IPO valuing it at over \$1 trillion. *SUF*¶92. OpenAI has achieved this commercial success by selling access to a series of “Large Language Models” (“LLMs”) known as the “GPT-series” through ChatGPT, its Application Programming Interface (“API”), and through a suite of Microsoft products powered by OpenAI’s LLMs, including, as relevant here, Copilot. *SUF*¶104-120, 125.

III. OpenAI Downloaded Millions of Books Without Payment.

A. OpenAI Torrented More Than 4 Million Books From The Pirate Site LibGen.

In October and November of 2018, OpenAI employee Alec Radford downloaded approximately 117,500 books from a pirate site called Library Genesis. The United States Trade Representative has identified LibGen as a “Notorious Market for Counterfeiting and Piracy” starting in 2017. *SUF*¶133. The site has been repeatedly shut down by courts and law enforcement for copyright infringement. *See, e.g., Elsevier Inc. v. Sci-Hub*, No. 15-cv-4282, 2017 WL 3868800, at *1 (S.D.N.Y. June 21, 2017); *SUF*¶135.

Like other LibGen users, Radford did not pay a purchase price for any of the books or acquire any license to use them. *SUF*¶153,177. Radford’s download included 96 of the Asserted Works. *SUF*¶159.

OpenAI kept these books in electronic storage, accessible to all company employees. *SUF*¶252-256. Around April 2019, OpenAI copied the Radford LibGen download to train a version of the LLM eventually known as GPT-3. *SUF*¶517. OpenAI made a considered decision to use books from LibGen, not license or purchase them. *SUF*¶161-181. Why? Because OpenAI

needed books quickly and cheaply; Microsoft wanted OpenAI to demonstrate its technology to Bill Gates in order to secure Microsoft's funding. SUF¶182-191, 519.

In June 2019, Microsoft infused OpenAI with \$1 billion under the companies' first Joint Development and Collaboration Agreement. SUF¶637-645. The purpose of the agreement was to commercialize OpenAI's LLMs; OpenAI formed a "capped-profit" entity for this purpose. SUF¶87. Though infused with cash after Microsoft's investment, OpenAI again decided against buying or licensing books. SUF¶192. Instead, it again turned to LibGen and downloaded millions more books free of charge. SUF¶193-243.

Between September 2019 and January 2020, OpenAI employees Benjamin Mann and Chris Hallacy torrented approximately 35 terabytes of data from LibGen. SUF¶193. The size of the Mann/Hallacy downloads matches the size of LibGen's entire fiction and nonfiction collections of 4.6 million books. SUF¶147,194. The Mann/Hallacy downloads included 187 of the Asserted Works. SUF¶934.

OpenAI employees understood at the time that they had sourced books from an illegal site. Sam Altman and Satya Nadella both testified that piracy of this kind is both "illegal" and "wrong." SUF¶370-372, 382-384. Internally, OpenAI employees commented that LibGen was made up of "pirated books," that it "violates copyright restrictions," and that it was "sketchy af [as f—k]." SUF¶352, 361, 366, 337-367. OpenAI took steps to conceal their piracy from the public. In internal notes from August 2019, an OpenAI employee wrote "[w]e trained GPT-3 on pirated stuff! No sharing that!" SUF¶344. When it published a research paper on the GPT-3 model, it changed references to book compilations internally dubbed "Libgen1" and "Libgen 2" to a more nondescript "Books1" and "Books2," a "deliberately vague" description "since it's libgen" and because of "fuzziness on licensing for this dataset." SUF¶269-282.

B. OpenAI Downloaded At Least 21,000 Books from the Pirated Books3 Dataset.

Books3 is a trove of 196,640 books sourced from a pirate site. SUF¶394-395. Books3 was included as a component of a larger compilation of datasets known as The Pile which was released on December 31, 2020. Shan Decl. Ex. A ¶85. While OpenAI has only produced a partial copy of the Books3 dataset, Plaintiffs’ expert has shown that OpenAI downloaded at least one full copy of Books3 containing 143 of the Asserted Works. SUF¶404, 411-417, 935. It is undisputed that even the partial copy of Books3 which OpenAI produced contained more than 20,000 books, including 14 of the Asserted Works. SUF¶935. As with LibGen, OpenAI employees understood that Books3 was illegal. SUF¶393-408.

C. OpenAI Downloaded Large Quantities of Books from Other Sources.

In addition to the millions of books OpenAI obtained by directly downloading from pirate sites, OpenAI also downloaded copies of books from targeted “scrapes” of the internet and third-party “crawling,” primarily from the organization “Common Crawl.” SUF¶425-447.

D. OpenAI And Microsoft Worked Together to Download Large Numbers of Books Via Crawling

OpenAI also sought data from Microsoft. SUF¶580-606. After Microsoft refused to allow OpenAI to use certain Microsoft crawled data, OpenAI decided in August and September 2023 to perform its own crawl of the internet. It did so using a program called “GPT-Bot.” SUF¶597, 602.

But OpenAI lacked the experience and infrastructure necessary for large-scale crawling. It needed Microsoft’s help and Microsoft obliged. SUF¶599. First, Microsoft transferred OpenAI a list of “seeding urls” which would guide OpenAI’s crawler. SUF¶598-599. Second, OpenAI signed a contract to pay Microsoft [REDACTED] so that Microsoft would crawl half of the web and transfer the downloaded content to OpenAI. SUF¶600. OpenAI would crawl the other half itself. SUF¶600. This secret initiative was dubbed “Project Mango.” SUF¶601, 606. The data

from this crawl was called the “GPT-bot” data and Plaintiffs’ expert has identified 49 of the Asserted Works in this data. *SUF*¶602, 939. As to the specific portion of the data downloaded by Microsoft and then transferred to OpenAI, Plaintiffs’ Expert has identified 27 copies of the Asserted Works in this data. *SUF*¶939.

IV. The OpenAI-Microsoft Partnership

A. The Joint Development and Collaboration Agreements.

The OpenAI-Microsoft partnership was memorialized primarily in three “Joint Development and Collaboration Agreements” (“JDCAs”) executed in 2019, 2021, and 2023. *See* *SUF*¶637, 646, 649.

The basic structure of the agreements had OpenAI share its AI technology with Microsoft, give Microsoft a profit and/or revenue share, collaborate with Microsoft in the design and building of bespoke supercomputers for OpenAI’s use, and commit to exclusive use of Microsoft’s cloud supercomputing platform (“Azure”). *See* *SUF*¶637-649, 698. In return, Microsoft provided OpenAI with a large investment [REDACTED] [REDACTED] and committed to designing and producing the hardware to serve OpenAI’s needs. *SUF*¶637, 638. Ultimately Microsoft invested a total of \$13 billion into OpenAI pursuant to the three JDCAs. *SUF*¶635.

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED] Nelson Decl.

Ex. 145 at -372; Ex. 146 at -613; Ex. 150 at -4102; *SUF*¶645-648, 690.

B. OpenAI and Microsoft Used Books as Currency in Their Commercial Relationship and Distributed Books to Each Other.

As noted above, OpenAI and Microsoft transferred large quantities of data to each other, as part of “Project Mango.” SUF¶¶600-604. They also distributed books to each other in at least three additional instances. SUF¶¶571-611. Defendants used copies of Plaintiffs’ works as leverage and currency in their business relationship, with one Microsoft employee characterizing the arrangement as “horse trading.” SUF¶¶589.

GPT-3 and LibGen Data: As part of a commercial agreement granting Microsoft access to GPT-3, OpenAI distributed copies of LibGen1 and LibGen2 to Microsoft in early 2021. SUF¶¶573-575. Microsoft downloaded and then later deleted the data. SUF¶¶577.

The Bing Index and Further GPT-3 Data Transfers: From the start of the OpenAI – Microsoft relationship, OpenAI wanted a copy of Microsoft’s “Bing Index.” SUF¶¶580. The Bing Index is an internet-wide crawl of the web that Microsoft created to use in its Bing search engine. SUF¶¶580.

Microsoft initially resisted due to copyright concerns. SUF¶¶585. Following extensive negotiations, and despite what they considered to be the [REDACTED] of the collaboration, the parties [REDACTED]

[REDACTED] SUF¶¶587, 594. As one OpenAI employee described it [REDACTED] [REDACTED].” SUF¶¶588.

Eventually, Microsoft transferred full copies of the Bing Index to OpenAI under the name “Project Taxi.” SUF¶¶593. These transfers included copies of Plaintiffs’ works which Microsoft had downloaded from the web through the Bing Crawl. *See* SUF¶¶581, 583-586, 594.

C. OpenAI Distributed LibGen Books To Third Party Contractors.

In addition to its distributions to Microsoft, OpenAI also distributed books to third-party contractors. *SUF*¶607-611. As part of a project regarding “book summarization,” OpenAI employees copied books from its LibGen sets and served them to contractors to read and summarize. *See id.*

V. OpenAI and Microsoft Exploited Plaintiffs’ Works to Train the At-Issue LLMs to Displace Human Authors.

A. The LLM Training Process and the In-Suit LLMs.

At-issue in this case are a number of OpenAI’s GPT-series of LLMs, specifically GPT-3 (released summer 2020); GPT-3.5 (2022); GPT-4 (2023); GPT-3.5-Turbo (2023); GPT-4 Turbo (2023); GPT-4o (2024); and GPT-4o mini (2024). *SUF*¶104. It is undisputed that the process of training these LLMs involves making numerous copies, including copies made when assembling and processing the training datasets, as well as copies of the training datasets made during the model training process. *SUF*¶535-47. And with the exception of GPT-4o mini, OpenAI copied Plaintiffs’ works into training datasets used in training all the at-issue LLMs. *SUF*¶93-104.

LLMs are a kind of “generative artificial intelligence” model, which respond to user inputs by generating outputs designed to emulate human expression. *SUF*¶94. The core of the LLM training process is called “pre-training,” and, while the parties dispute almost every aspect of this process, it is undisputed that the pre-training process involves adjusting the model via a process of guess-and-check. *SUF*¶95-96. In this process, training data, e.g. an incomplete sentence from a book, is fed into the model, the model outputs a prediction of what the next word should be, that prediction is compared against the actual continuation, e.g. the full sentence from the book, and the training code adjusts the model’s internal mathematical components so that the model will more reliably reproduce the “correct” answer from its training data. *SUF*¶96-97. These internal components are referred to as the model’s “weights” or “parameters.” *SUF*¶98. As one OpenAI

presentation described it “the training objective” made its “models want to regurgitate” the text on which they were trained. *SUF*¶101.

The data used to train an LLM has a major impact on a model’s capabilities. *SUF*¶102-103, 451. While it is possible to develop an LLM on any large corpus of text, including public domain works, *SUF*¶301-03, better quality writing helped the model mimic better-quality human expression—or, as OpenAI competitor Anthropic put it—“writing that an editor would approve of,” Zhao Decl. Ex. A ¶48. Books, for that reason, are high-quality data because they contain organized writing, structured thoughts, and, as former OpenAI engineer Ben Mann agreed, “coherent storytelling.” *SUF*¶471. During the training process, the LLM copies and encodes the works’ expressive elements (like syntax, sentence structure, and word choice and ordering) into the machine, enabling it to better mimic the qualities of the works on which it was trained. *SUF*¶101.

B. OpenAI Copied Plaintiffs’ Books to Train In-Suit LLMs.

LibGen1 and LibGen2 comprised roughly 570,000 books from the LibGen downloads and were copied repeatedly for LLM training. *SUF*¶138-81, 202, 192-243. Versions of these datasets were used to train GPT-3 and GPT-3.5, the model that powered ChatGPT’s viral launch in November 2022. *SUF*¶519. Plaintiffs’ expert has identified 124 of the Asserted Works in the two LibGen compilations. *SUF*¶936.

OpenAI deleted its LibGen files in the summer of 2022 due to legal concerns. *SUF*¶283-328. Prior to their deletion, all the LibGen books downloaded by Mann, Hallacy, and Radford, as well as the two LibGen training datasets were, per OpenAI’s data access policies, available to OpenAI’s employees and contractors to use as they saw fit. *SUF*¶252-56. As part of the records created during the deletion—known internally as “Project Clear”—OpenAI compiled a list of all internal copies and versions LibGen datasets. *SUF*¶314, 320. The list numbered near 500 distinct

entries and encompassed a variety of uses at the time of deletion. *SUF*¶320. The LibGen book compilations are the only two training corpuses OpenAI has ever deleted. *SUF*¶324. As one employee characterized reports of similar conduct by a competitor, the plan was to “[t]rain on libgen, raise money, get new data, delete the old models, be clean forever more, temporal regulatory arbitrage.” *SUF*¶362. Other employees reacted to the message with “:not-legal-but-very-cool:.” *SUF*¶362.

OpenAI similarly made copies of the books present in data downloaded from, *inter alia*, scrapes, Common Crawl, and the GPT-Bot/Mango crawl to create datasets used in training the at-issue LLMs. *SUF*¶537. Plaintiffs’ expert has identified 157 of the Asserted Works in the data OpenAI used to train the in-suit models. *SUF*¶937.

C. OpenAI and Microsoft Valued and Licensed Books for Their Expressive Qualities.

Throughout OpenAI’s acquisition of books, OpenAI and Microsoft executives and employees consistently stated their belief that books were valuable because, while not required to train an LLM, they improved LLM performance (particularly on generating competing writing) and reduced training costs. *SUF*¶297, 301-03, 467-85, 706-13, 745-49, 784. Both OpenAI and Microsoft executives and employees expressed their willingness to pay substantial sums—including “10X” of the hundreds of millions the companies spent on “training a model”—to acquire and license books because “you can reuse the data you acquired.” *SUF*¶297; *see* *SUF*¶789. High-quality copyrighted books enabled OpenAI to train its LLMs faster and with less compute, saving OpenAI substantial sums. *SUF*¶485.

Because of books’ acknowledged value, both OpenAI and Microsoft have paid for licenses to books for AI training, *SUF*¶775-814, including, in Microsoft’s case [REDACTED], *SUF*¶801.

OpenAI and Microsoft are not alone, as across the industry AI companies are paying for training licenses or for past use in AI training. SUF¶825-36.

D. The In-Suit LLMs Reproduce Plaintiffs’ Copyrighted Material to End Users.

As a result of training on Plaintiffs’ copyrighted expression, OpenAI’s LLMs have the capability to, and in fact do, regurgitate portions of Plaintiffs’ works to end users when prompted. SUF¶548-70. While OpenAI eventually attempted to mitigate this issue, it launched GPT-3 and GPT-3.5 (and ChatGPT) without any anti-regurgitation protections. SUF¶564. It is similarly undisputed that users have sought to use the models to write sequels to Plaintiffs’ works. SUF¶736-39.

E. OpenAI Designed Its LLMs to Replace Human Creations.

OpenAI employees understood that their models’ performance was not defined by the technical choices OpenAI made regarding model design but rather was “determined by your dataset, nothing else.” SUF¶103.

From its earliest LLMs, OpenAI trained on books because they enabled the model to output text that resembled books. SUF¶448-53. OpenAI’s first LLM, GPT-1, trained exclusively on 7,000 books, SUF¶449, and, only a few months after its release, Alec Radford went to LibGen to download over 117,000 more, SUF¶149. OpenAI employees repeatedly expressed their views that the models were excellent at writing because they trained on books. SUF¶709-17. When OpenAI wanted to improve writing performance, the default was to train the models more on “books.” SUF¶713.

The ability of its models to write like a human author has only grown more important to OpenAI’s bottom-line over time. [REDACTED]

[REDACTED] SUF¶723. [REDACTED]

████████████████████ And both Microsoft and OpenAI have extensively marketed the use of the in-suit LLMs for writing. SUF¶740-44.

OpenAI hired Tarun Gogineni in 2022 to head OpenAI’s efforts to improve model writing quality. SUF¶714. In his role, Gogineni described himself as having “an undue amount of power over the english language,” SUF¶716, and stated his belief that, starting with GPT-3, OpenAI’s models were already “very very good at writing stories” because they “had such a good understanding of writing style,” SUF¶717. Gogineni knew the models he was training would displace authors, but thought that such displacement constituted “acceptable economic disruption.” SUF¶731.

In 2025, nearly two years after George R. R. Martin had sued OpenAI in this lawsuit, Gogineni tweeted that his “research mission” was to have GPT models write the “last two books of [Martin’s] *A Song of Ice and Fire*” series and that OpenAI would “pour several exaflop [of computer power] into finishing it one day.” SUF¶727-28. Gogineni similarly tweeted that he “rest[s] easy knowing that even if GRRM [George R. R. Martin] dies early, GPT-5 will autocomplete his series,” SUF¶728, and that OpenAI should continue developing its LLMs’ capabilities “so GPT can write the next *[S]ong of [I]ce and [F]ire* book,” SUF¶728.

Gogineni was not alone. One novelist employed by OpenAI said he worked there because his “days [were] numbered” as soon the models would “write the books!” SUF¶711. Jack Clark, OpenAI’s former policy director, was clear in internal memos that “our work on AI and Creativity is going to increasingly lead to us creating systems that substitute for the labor of [] people . . . *The better we do on GPT-X, the more worried genre fiction authors will become about us substituting for them on Amazon . . . Our work in this area will make people unemployed . . . There will be a point where a bunch of artists express worry about what we’re*

doing here and we'll likely ignore their concerns and release anyway . . ." SUF¶749 (emphases added). As he put it, "[t]he blunt truth: OpenAI is an entity of capital, with a capital-oriented mindset around its end-product (AGI)." SUF¶750. OpenAI's GPT models, Clark wrote, "*ultimately substitute[] for the practiced, deeply personal labor of the creative communities of the world.*" SUF¶750 (emphasis added). In contrast to AI which "invent[s] entirely new things," OpenAI's models "synthesize[d] stuff that will seem *substitutive* for the personal labor of people." SUF¶750 (emphasis added).

Microsoft scientists came to the same conclusion. Brent Hecht, Microsoft's Director of Applied Science, wrote that Microsoft's failure to license data for training "has started a doom loop" because the LLM "end product threatens the economic foundations of its essential suppliers," that is, human creators. SUF¶765. Because LLMs "replicate the patterns in the training content," failure to "distribut[e] economic value down the supply chain, necessarily threatens the economic sustainability of those who create the content." SUF¶765. Hecht testified that he had serious concern that book authors would suffer "economic harm and perhaps not be able to continue in their business[]" due to the outputs of LLMs trained on their work without compensation. SUF¶767.

F. [REDACTED] AI-Generated Works Are Rapidly Entering the Market and Gaining Sales.

What OpenAI and Microsoft predicted is reflected in the marketplace. [REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]. SUF¶851; SUF¶849-54. [REDACTED]

[REDACTED]. SUF¶851. [REDACTED]

[REDACTED]

conclusory statements or on contentions that the affidavits supporting the motion are not credible.”
Id.

II. Copyright Law

A. Direct Infringement

Establishing a *prima facie* case of direct copyright infringement requires “(1) ownership of a valid copyright, and (2) copying of constituent elements of the work that are original.” *Williams v. Crichton*, 84 F.3d 581, 587 (2d Cir. 1996). “A certificate of registration from the United States Register of Copyrights constitutes prima facie evidence of the valid ownership of a copyright.” *Jorgensen v. Epic/Sony Records*, 351 F.3d 46, 51 (2d Cir. 2003). “[C]opying’ is shorthand for the infringing of any of the copyright owner’s five exclusive rights described in [17 U.S.C.] § 106.” *Arista Records, LLC v. Doe 3*, 604 F.3d 110, 117 (2d Cir. 2010). Those rights include the right of reproduction and distribution. *Id.* “The legal or beneficial owner of an exclusive right under a copyright is entitled . . . to institute an action for any infringement of that particular right committed while he or she is the owner of it.” 17 U.S.C.A. § 501(b)).

B. Vicarious Infringement

Vicarious infringement occurs where a defendant (1) “had the right and ability to supervise” the infringing conduct and (2) a “direct financial interest” in the same. *See EMI Christian Music Grp., Inc. v. MP3tunes, LLC*, 844 F.3d 79, 99 (2d Cir. 2016).

As to the first element, “a defendant need not have ‘formal power to control’ where a direct infringer ‘depend[s] upon [the defendant] for direction.’ . . . Rather, ‘[t]he ability to block infringers’ access to a particular environment for any reason whatsoever is evidence of the right and ability to supervise.’” *Arista Recs. LLC v. Usenet.com, Inc.*, 633 F. Supp. 2d 124, 157 (S.D.N.Y. 2009) (cleaned up). “If a defendant has the right to block access, such right must be ‘exercised to its fullest extent’ to ‘escape imposition of vicarious liability.’” *In re Frontier*

Communications Corporation, 658 B.R. 277, 293 (Bankr. S.D.N.Y. 2024) (quoting *Napster*, 239 F.3d at 1023).

C. Fair Use

Because fair use is an affirmative defense, Defendants bear the burden of proof and Plaintiffs are entitled to summary judgment when they show that there is an absence of “evidence to support an essential element of the non-moving party’s case.” *O’Neil v. Ratajkowski*, 563 F. Supp. 3d 112, 128 (S.D.N.Y. 2021) (quotation omitted). In determining whether a particular use is fair, courts consider four factors: (1) “the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes”; (2) “the nature of the copyrighted work”; (3) “the amount and substantiality of the portion used in relation to the copyrighted work as a whole”; and (4) “the effect of the use upon the potential market for or value of the copyrighted work.” 17 U.S.C. § 107. The fourth factor is “undoubtedly the single most important element of fair use.” *Hachette Book Grp., Inc. v. Internet Archive*, 115 F.4th 163, 189 (2d Cir. 2024).

ARGUMENT

I. The Elements of a Copyright Infringement Claim Are Satisfied As a Matter of Law.

A. Plaintiffs Own Valid Copyrights in Their Works.

Plaintiffs are the owners of the copyrights in the registered Asserted Works. SUF¶3. Because all Plaintiffs receive royalties for their Works, they are at a minimum beneficial owners. SUF¶5-67; *see Harris v. Simon & Schuster, Inc.*, 646 F. Supp. 2d 622, 632 (S.D.N.Y. 2009).

B. OpenAI and Microsoft Copied Plaintiffs’ Works When Initially Downloading the Works, When Training Their LLMs, and When Distributing the Works To One Another.

OpenAI and Microsoft both copied Plaintiffs’ works repeatedly, in at least three separate ways. *First*, OpenAI reproduced Plaintiffs’ works when downloading them from LibGen and

Books3. *Supra* Factual Background § III. *Second*, OpenAI made additional reproductions of Plaintiffs’ books in training its LLMs. *Supra id.* *Third*, both OpenAI and Microsoft distributed copies of Plaintiffs works. *Supra* § IV.B. Each of these sets of actions constitutes infringement of Plaintiffs’ exclusive rights. *See, e.g., Arista Recs. LLC v. Usenet.com, Inc.*, 633 F. Supp. 2d 124, 149-50 (S.D.N.Y. 2009).

II. OpenAI’s Unauthorized Downloading of Plaintiffs’ Works from LibGen and Books3 Was Not Fair Use as a Matter of Law.

The first “use” on which Plaintiffs are entitled to summary judgment is OpenAI’s unauthorized acquisition of their copyrighted books by downloading from LibGen and via downloading the Books3 dataset. Under *Warhol*, a court must consider each use separately, 598 U.S. at 533. “Pirating copies to build a research library without paying for it, and to retain copies should they prove useful for one thing or another, was its own use—and not a transformative one.” *Bartz*, 787 F. Supp. 3d at 1026 (citations omitted).

Courts have consistently held that downloading copyrighted works for free to avoid paying for them is a manifestly unfair use. *See Napster*, 239 F.3d at 1019; *BMG Music v. Gonzalez*, 430 F.3d 888, 891 (7th Cir. 2005) (“[D]ownloading full copies of copyrighted material without compensation to authors cannot be deemed ‘fair use’”). There is no exception to this general rule for large technology companies downloading books for profit. The opposite: All else equal, the commercial exploitation of copyrighted works weighs *against* fair use. *See Fox News Network, LLC v. Tveyes, Inc.*, 883 F.3d 169, 178 (2d Cir. 2018). For the reasons set forth below, each fair use factor weighs against OpenAI’s downloading as a matter of law.

A. Factor One: OpenAI’s Downloading from Pirate Sites Was Nontransformative, Lacks Justification, and Is Blatantly Commercial.

The first factor is the purpose and character of the use. 17 U.S.C. § 107(1). Under this factor, courts consider whether a defendant had a sufficient “independent justification” for a use,

Warhol, 598 U.S. at 547; whether and the degree to which a use is “transformative”—that is, whether it “adds something new, with a further purpose or different character, altering the first with new expression, meaning or message,” *Campbell v. Acuff-Rose Music Inc.*, 510 U.S. 569, 579 (1994); and “the commercial nature of the use,” *Warhol*, 598 U.S. at 532–33. Courts have consistently held that the first factor weighs against fair use where a copyist “profit[s] from the exploitation of the copyrighted work without paying the customary prices.” *Arista Recs., Inc. v. Mp3Board, Inc.*, 2002 WL 1997918, at *13 (S.D.N.Y. Aug. 29, 2002) (Stein, J.) (quoting *Harper & Row*, 471 U.S. at 562); *Napster*, 239 F.3d at 1015 (rejecting fair use defense for Napster users who “get for free something they would ordinarily have to buy”); *BMG Music*, 430 F.3d at 890 (“[a] copy downloaded, played, and retained on one’s hard drive for future use is a direct substitute for a purchased copy”).

OpenAI has no hope of success on the first factor. Like teenagers in the 2000s, OpenAI downloaded books for free while knowing it could have purchased or licensed copies precisely to avoid paying for them, substituting books sales for free downloads. *See Internet Archive*, 115 F. 4th at 190 (free digital book copies “usurp the market for the originals” because it is “difficult to compete with free”); *SUF* ¶¶158-181, 283-304. OpenAI’s “repeated and exploitative unauthorized copies of copyrighted works were made to save the expense of purchasing authorized copies.” *Napster*, 239 F.3d at 1015. The principal difference between OpenAI and the adolescent pirates of yesteryear is that OpenAI’s conduct was far more commercial—further weakening an already meritless defense. *See TVEyes*, 883 F.3d at 178.

OpenAI will claim that their unauthorized downloading is protected fair use because, *after* it downloaded pirated copies, it used some for a supposedly “transformative” use—LLM training and “research.” That is wrong. For starters, the supposedly “transformative” use was not

transformative at all, *see infra* § III.A. More importantly, however, an overarching “fair use” purpose does not inoculate every act of copying. As *Warhol* recognizes, a court must analyze each separate use separately—just as each quotation in a book review must be analyzed separately, OpenAI’s downloading must be analyzed separately from its later use. *See Warhol*, 598 U.S. at 532-33 & n.7. *Bartz* confronted this exact question in a case involving the same pirate libraries at issue here. As Judge Alsup wrote, “the person who copies” a book from “a pirate site has infringed already, full stop Such piracy of otherwise available copies is inherently, irredeemably infringing even if the pirated copies are immediately used for the transformative use and immediately discarded.” 787 F. Supp. 3d at 1025.

The Second Circuit has likewise repeatedly rejected claims that a downstream use can be used to bless other acts of copying. In *American Geophysical Union v. Texaco Inc.*, the Second Circuit explained that a company could not make unlimited photocopies of copyrighted articles, even if all the copies were ultimately in pursuit of “research.” 60 F.3d at 924. Under such a rule, “the concept of a ‘transformative’ use would be extended beyond recognition.” *Id.* *Texaco* is unequivocal: “Texaco cannot gain fair use insulation for [its employee’s] archival photocopying of articles (or books) simply because such copying is done by a company doing research.” *Id.* The Court continued: “It would be equally extravagant for a newspaper to contend that because its business is ‘news reporting’ it may line the shelves of its reporters with photocopies of books on journalism.” *Id.*; *see also TVEyes, Inc.*, 883 F.3d at 178 n.4 (“[t]hat a secondary use can facilitate research does not itself support a finding that the secondary use is transformative.”); *Basic Books, Inc. v. Kinko’s Graphics Corp.*, 758 F. Supp. 1522, 1531 (S.D.N.Y. 1991) (no fair use defense even where “use of the Kinko’s packets, in the hands of [] students [to which Kinko’s supplied the copies], was no doubt educational”). If purposes expressly enshrined in Section 107’s preamble

like research and news reporting were insufficient to inoculate a copyist from an otherwise unauthorized acquisition, Defendants’ claimed “fair use” purposes fare no better.

Lastly, a number of books OpenAI downloaded were *never* used for LLM training but instead used to stock OpenAI’s central library. *Supra* Factual Background §§ III.A V.B. OpenAI’s data access policies make clear that all “full-time employees” and contractors had full access to the books for whatever purpose they wished, including to distribute them to third-parties. *See* SUP¶254. And under *Texaco*, such archival copying—the retention and making available internally of millions of books to its employees for whatever purpose they wished to put them to—is nontransformative and not fair use. *See* 60 F.3d at 924 (discussing hypothetical where company “sent around entire books” to “researchers” to illustrate non-transformative “archival purposes”). As Judge Alsup put it when considering identical conduct, “piracy was the point: To build a central library that one could have paid for . . . but without paying for it.” *Bartz*, 787 F. Supp. 3d at 1027.³

B. Factor Two: Plaintiffs’ Copyrighted Books are at the Core of Copyright Protection.

The second fair use factor “calls for recognition that some works are closer to the core of intended copyright protection than others, with the consequence that fair use is more difficult to establish when the former works are copied.” *Campbell*, 510 U.S. at 586. “[B]oth fiction and nonfiction books,” like those by the Plaintiffs here, “contain original expression close to the core of intended copyright protection.” *Internet Archive*, 115 F.4th at 187 (internal quotation omitted). And that is especially true when the use in question—here a full scale reproduction to avoid paying

³ This also distinguishes *Kadrey*, as there the court found that the sole purpose of Meta’s downloading was for LLM training. 788, F. Supp. 3d at 1048. But here OpenAI here acquired Plaintiffs’ works and put them to multiple other uses, including by distributing them to Microsoft and third-parties. *See supra* Factual Background § IV.B.

for a legitimate copy—is non-transformative. *See Warner Bros. Ent. Inc. v. RDR Books*, 575 F. Supp. 2d 513, 549 (S.D.N.Y. 2008). This factor thus weighs against fair use.

C. Factor Three: OpenAI Pirated Plaintiffs’ Books in Their Entirety.

Under the third factor, courts consider whether “the amount and substantiality of the portion used in relation to the copyrighted work as a whole . . . are reasonable in relation to the purpose of the copying.” *Campbell*, 510 U.S. at 586–87. “[C]opying an entire work militates against a finding of fair use,” and that is what OpenAI indisputedly did here. *Napster*, 239 F.3d at 1016 (citation omitted).

D. Factor Four: Piracy Harms the Market for Plaintiffs’ Copyrighted Books

The fourth and final factor is the effect of the copying on the “potential market for or value of the copyrighted work.” 17 U.S.C. § 107(4). This factor encompasses “not only the market harm caused by the particular actions of the alleged infringer, but also the market harm that would result from *unrestricted and widespread conduct* of the same sort.” *Internet Archive*, 115 F.4th at 189 (quoting *TVEyes*, 883 F.3d at 179) (emphasis added). In other words, this factor considers both the actual market harm caused by the defendant’s actions *and* the potential market harm that would occur if everyone engaged in the same conduct. The fourth factor is “undoubtedly the single most important element of fair use.” *Id.* at 189 (quoting *Harper & Row*, 471 U.S. at 566). This factor weighs strongly against fair use for multiple independent reasons.

First, OpenAI’s unauthorized, unpaid, downloads directly harmed the markets for Plaintiffs’ books by obtaining the “same product” that traditional, lawful, purchasers obtain, “but at no cost.” *Internet Archive*, 115 F.4th at 190. “Each digital book” OpenAI copied for free “deprives [] authors of the revenues due to them as compensation for their unique creations.” *Id.* at 195; *see Byrne v. Brit. Broad. Corp.*, 132 F. Supp. 2d 229, 236 (S.D.N.Y. 2001) (Stein, J.) (when a user makes an “unauthorized copy” of a work, it “has no need to purchase a copy from the

copyright holder or his licensees, thus lessening the market” for the work); *BMG Music*, 430 F.3d at 890–91 (“[A]ll 1,000+ of [defendant’s] downloads . . . undermined the means by which authors seek to profit.”). OpenAI’s copies thus “usurp the market for the originals” because it is “difficult to compete with free.” *Internet Archive*, 115 F. 4th at 190 (internal quotations omitted).

Second, as the Second Circuit held in the context of free distribution of ebook copies in *Internet Archive*, the result of “unrestricted and widespread” conduct like OpenAI’s unpaid downloads from pirate sites would be the “decimat[ion] [of] markets for the Works in Suit across formats.” 115 F.4th at 195. If anyone could lawfully obtain any book they wanted for free, “there would be little reason . . . to pay [] for content.” *Id.* at 193; *see Arista Recs.*, 604 F.3d at 124 (“the likely detrimental effect of file-sharing on the value of copyrighted compositions is well-documented” (citation omitted)).

The record here demonstrates the harms of OpenAI’s piracy clearly. As Plaintiffs’ expert Mark Seeley explains, piracy annually costs authors hundreds of millions of dollars in lost sales. *See* Seeley Decl. Ex. A at ¶53-62. And OpenAI’s piracy had the additional effect of harming authors by “normaliz[ing] piracy” throughout the AI industry and “encourag[ing] others to use shadow libraries.” *Id.* at § VIII, ¶63. Shadow libraries themselves have reported that, prior to AI, “shadow-libraries were dying,” but then were revitalized when “[v]irtually all major companies building LLMs contacted us to train on our data.” *Id.* at ¶65, 63-76. Moreover, as Plaintiffs expert Dr. David Choffnes demonstrated, when OpenAI torrented LibGen, it used a setting which directly benefitted other users downloading files from LibGen. *See* Choffnes Decl. Ex. B at ¶33-43.

That is dispositive of OpenAI’s fair use defense as a matter of law: “to negate fair use one need only show that if the challenged use ‘should become widespread, it would adversely affect the *potential* market for the copyrighted work.’” *Harper & Row Publishers, Inc. v. Nation Enters.*,

471 U.S. 539, 568 (1985) (quotation omitted). As the *Internet Archive* court held, even in the absence of “empirical data” regarding the impact of such widespread use, the court may “rely on” the simple “logical inference[]” that widespread free acquisition of the works would harm the market for the works. *Internet Archive*, 115 F.4th at 193. It is “self-evident that if [OpenAI’s] use were to become widespread, it would adversely affect [the] markets for the Works in Suit.” *Id.* (quotation omitted); *accord Byrne*, 132 F. Supp. 2d at 236 (widespread uncompensated copying would “lessen[] the market for the song”).

III. Defendants’ Copying of Plaintiffs’ Works to Train LLMs Was Not Fair Use.

Copyright “reflect[s] a considered judgment that ‘the Progress of Science and useful Arts’ is best promoted by laws that protect authors’ original works and permit authors to set the terms of engagement, at least for a limited time.” *Internet Archive*, 115 F.4th at 195 (quoting U.S. Const. Art. I § 8, cl. 8). OpenAI’s unlicensed copying of Plaintiffs’ books—copying both OpenAI and Microsoft admit was unnecessary to any purportedly transformative purpose—in order to train machines to imitate and, ultimately, supplant Plaintiffs’ works threatens to decimate the precise incentives copyright exists to preserve.

A. Factor One: OpenAI’s Copying of Plaintiffs’ Books To Develop LLMs Was Non-Transformative, Lacks Any Independent Justification, and Was Commercial.

Under factor one—the “purpose and character” of the use—courts must weigh the “transformativeness” of the use, the commerciality of the use, and the objective “justification” connecting the copying to the purportedly transformative purpose. *See supra* § II.A. The purpose and character of OpenAI’s copying of Plaintiffs’ works for use in LLM training weighs strongly against fair use on each count. *First*, OpenAI’s exploitation was nontransformative because it exploited the protected expression in Plaintiffs’ works for the purpose of developing a competing commercial product. *Second*, OpenAI (and Microsoft) admit that their use of Plaintiffs’ works was

not necessary for any allegedly transformative purpose of developing a general purpose LLM. Instead, Plaintiffs’ works were valuable in achieving the substitutive end of supplanting the work of Plaintiffs’ and authors more generally. *Third*, OpenAI’s use was commercial. This is, as *Warhol* put it, a case where all the first factor’s “elements point in the same direction”—against fair use. 598 U.S. at 538.

i) OpenAI Copied Plaintiffs’ Works To Develop A Product Designed To Generate Competing Output.

The Supreme Court has cautioned against an “overbroad concept of transformative use”: A use does not become transformative simply by adding “any further purpose” or “different character,” but instead the Court must consider the “degree of transformation.” *Warhol*, 598 U.S. at 529. The central question is “the problem of substitution—copyright’s *bête noire*.” *Id.* at 528. Here, OpenAI cannot meet its burden to establish that its exploitation of Plaintiffs’ works was transformative.

First, OpenAI and Microsoft’s internal documents, testimony, and public statements all confirm that OpenAI’s exploitation of Plaintiffs’ works to build its GPT models “run[s] the risk of substitution” for human-authored books. *Id.* at 532. OpenAI employees recognized that their models were “very very good at writing stories,” SUF¶604, could “mimic the style of a specific writer,” SUF¶588, would generate “economic disruption” as a result, SUF¶610, and that novelists’ “days were numbered,” SUF¶597. Both OpenAI and Microsoft executives recognized that their exploitation posed an existential threat to book authors like Plaintiffs: OpenAI’s Clark recognized that the GPT models would “ultimately substitute[] for the practiced, deeply personal labor of the creative communities of the world,” while Microsoft’s Hecht warned of the “doom loop” that would decimate the market for human-authored books. SUF¶749-68. Because OpenAI’s copying was intended to “make it easier to develop [] competing” products—namely, a machine capable

of generating writing that would supplant Plaintiffs’ works in the marketplace—this is precisely the type of use that “is not transformative” as a matter of law. *Thomson Reuters Enter. Ctr. GMBH v. Ross Intel. Inc.*, 765 F. Supp. 3d 382, 397–99 (D. Del. 2025) (granting summary judgment).

Second, OpenAI’s copying is nontransformative because the training process and resulting model exploits precisely what the Copyright Act protects—the protected expression in Plaintiffs’ works—in order to generate competing works. Both OpenAI and Microsoft’s experts admit that a model’s ability to generate “book-like” text derives from the training on books. *See, e.g.*, SUF¶745-46; SUF¶748 (“book-like text” can improve the “model’s ability to generate [] book-like outputs.”); *see also, e.g.*, Zhao Decl. Ex. B ¶50 (“a model trained on copyrighted books is optimizing to encode and reproduce the author’s expressive choices”). As one OpenAI engineer put it, what a model generates is “determined by [its] dataset, nothing else.” SUF¶103. The “dataset,” of course, is the copyrighted expression—including Plaintiffs’ works—that OpenAI copied repeatedly to develop its GPT models. This reflects the basic principle of garbage-in-garbage-out: The better quality prose used to train the model, the more likely the model will be capable of generating prose that can compete with the copyrighted material on which it was trained.

This places OpenAI’s copying here miles away from cases like *Google Books*, 804 F.3d 202, and *Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87 (2d Cir. 2014)—neither of which involved conduct designed to exploit authors’ protected expression to build a substitutive commercial product. *Google Books*, which “test[ed] the boundaries of fair use,” conveyed to the public information *about* books by identifying books that hit on particular keywords searches. 804 F.3d at 206, 214-20. The functions at issue gave users no “meaningful experience of the expressive content of the book,” and thus posed no risk of serving as a competing substitute. *Id.* at 227. The same was true in *HathiTrust*. 755 F.3d at 97-102. This case is not about copying novels to create

an index: OpenAI copied Plaintiffs’ works to extract the plot, prose, and other protected elements, all in order to manufacture a machine capable of flooding the market with competing works. No decision of this court has called that fair use.

While both the *Kadrey* and *Bartz* decisions concluded that the use of books to train other LLMs was transformative, both did so on records lacking the evidence of substitutive purpose here. In *Kadrey*, although the court stated that “in most cases” AI training would not be fair use due to the “potential to flood the market with endless amounts of . . . books,” the court found that factor one favored Meta’s use of books in training its LLMs because there was no “indication that . . . Meta developed [its LLM] with the *purpose* of enabling it to create books that compete with the plaintiffs’[.]” 788 F. Supp. 3d at 1034, 1045 (emphasis added). Similarly *Bartz* found that the first factor favored Anthropic because “Anthropic’s LLMs trained upon works not to . . . *supplant* them.” 787 F. Supp. 3d at 1022 (emphasis added).

Here, the record is replete with evidence that OpenAI developed its LLMs with the purpose of creating competing books and indeed optimized its LLMs to “supplant” Plaintiffs in this case. *See supra* Factual Background § V.A, E. Among other things, OpenAI’s internal documents confirm that its models would “substitute” for the “labor” of creators, and OpenAI’s head of writing quality repeatedly affirmed his intent to have the models substitute for authors *in this case*. *SUF* ¶749. As *Kadrey* itself acknowledged, factor one would favor plaintiffs if an LLM was “designed to be used to create works substantially similar to those on which it was trained, or to create works that competed with the originals without being substantially similar.” 788 F. Supp. 3d at 1045 n.6. That is precisely what occurred here.

To the extent that either decision can be read to hold that LLM training is a transformative use because LLM *technology* is transformative, that cannot be squared with the law of this Circuit.

See HathiTrust, 755 F.3d at 96 (“[A] use does not become transformative by making an invaluable contribution to the progress of science and cultivation of the arts.” (quotation omitted)). Instead, the “central question” under factor one is substitution, *Warhol*, 598 U.S. at 528, and here OpenAI’s copying of Plaintiffs’ works was, in OpenAI’s own words, designed to create a machine that would “substitute” for Plaintiffs’ “creative labor.” *See supra* Factual Background § V.E; SUP¶749-50.

Because the use here improved the model’s capabilities to generate an infinite flood of “competing” works to substitute for Plaintiffs’ works (and for Defendants’ gain), the copying is “not transformative.” *Ross Intel. Inc.*, 765 F. Supp. 3d at 399.

ii) *OpenAI’s Copying of Plaintiffs’ Works Was Not “Reasonably Necessary” To Achieve Its Purportedly Transformative Purpose.*

Even if the Court were to find that OpenAI’s copying of Plaintiffs’ works to train its LLMs was transformative, factor one would still weigh heavily against its use because OpenAI (and Microsoft) admit they had no justification for their copying.

Because “fair use is an affirmative defense,” Defendants “bear[] the burden to *justify* [their] taking of [Plaintiffs’ works] with some reason other than ‘I can make it better.’” *Warhol*, 598 U.S. at 547 n.21 (emphasis added); *see Romanova v. Amilus Inc.*, 138 F.4th 104, 112 (2d Cir. 2025) (“*Warhol* made clear that *Campbell*’s requirement of justification is not applicable only in cases of claimed parody, but applies generally to all claims of fair use.”).

As *Warhol* explains, the question under factor one relates to the nexus—the “justification”—between the copying in question and the purportedly transformative purpose. 598 U.S. at 547. Copying can be “justified” where it’s “reasonably necessary to achieve the user’s new purpose” (e.g. a parodist’s “need[] to mimic an original to make its point”). *Id.* at 532. But an “independent justification” is particularly needed where, as here, “an original work and copying use share the same or highly similar purposes, or *where wide dissemination of a secondary work*

would otherwise run the risk of substitution for the original or licensed derivatives of it.” *Id.* at 532 (emphasis added).

For example, in *Warhol* itself, even though the reproduction of the plaintiff’s photograph in an Andy Warhol canvas *may* have been a transformative fair use of the photograph as a piece of art commenting on celebrity, there was no justification for why the *Warhol* defendant *needed* to reproduce that *particular* photograph in a magazine which connected that reproduction to the transformative aspects of the underlying portrait. *See, e.g.*, 598 U.S. at 539 & n.15 (contrasting Warhol’s use of the “Campbell’s Soup” logo in an art display versus the use of Warhol’s art by a “Campbell’s Soup” competitor); *see also id.* at 545-48; *McGucken v. Pub Ocean Ltd. et al.*, 42 F.4th 1149, 1158-61 (9th Cir. 2022) (finding first factor favored copyright owner where use of photo to illustrate article had no connection to transformative purpose).

Defendants cannot establish a justification for their copying on this record because they admit copying Plaintiffs’ works was not *necessary* to develop the purportedly “transformative” character of the at-issue LLMs: their ability to “generalize” and be put to general-purpose uses. Defendants acknowledge that they could develop a functional LLM—one that, perhaps, provides recommendations on the best nearby park, or writes computer code—without training on, e.g., *Game of Thrones* or *Cleopatra: A Life*.

Defendants’ lack of any cognizable justification under the first factor is not a matter of factual dispute. OpenAI’s technical expert, Dr. Chris Callison-Burch opined that the Asserted Works were “not technically required” for OpenAI’s use; he agrees that no individual work or category of works (such as books) was necessary for model performance; and he concludes that including or excluding any specific book or group of books would have no meaningful effect on model performance. *See* SUPP ¶¶496-501. As he put it, “[t]here is no reason to believe that the

Asserted Works individually had any meaningful impact on model performance . . . no individual work was necessary for inclusion in the corpora.” SUF¶500. Microsoft’s technical expert Dr. John Lafferty likewise opined that no individual work, subset, collection, or even category of texts (such as Plaintiffs’ books or books in general) was necessary to confer the general-purpose capabilities at issue. *See* SUF¶502-510. And both OpenAI and Microsoft witnesses testified that they understood the books were not required to train their LLMs or necessary to the purportedly transformative aspects of “generalization.” SUF¶511-16. Instead, the record shows that OpenAI sought out books because they saved money on LLM training and improved the models’ performance at producing competing works. *See supra* Factual Background § V.C.

That is not enough. Copying a work “often is” “helpful,” *Warhol*, 598 U.S. at 547, but OpenAI must show that the copying itself played an important role in the achievement of its transformative purpose, as in the case of parody, *id.* at 532.

This is a longstanding principle. In *Byrne*, this Court addressed a fair use defense by the BBC, which had recorded a copyrighted song to soundtrack portions of a broadcast on Irish politics. 132 F. Supp. 2d at 231-32. This court rejected the contention that the BBC’s self-applied label of “news reporting” controlled. *Id.* at 234. Underscoring the extent of the justification required, the court stated that “even if the Song were recorded as part of an effort” at news reporting, the existence of “*equally informative non-infringing alternatives*” would preclude a finding for the BBC on factor one. *Id.* at 235 (emphasis added).

Here, it is undisputed that OpenAI could have used other data (including the “equally informative non-infringing alternative[]” of public domain books) and still produced functional and performant LLMs. *See* SUF¶301-03. Indeed, at least one in-suit LLM—GPT-4o-mini—did not train on any of Plaintiffs’ works. SUF¶533-34. Even if the Court were to find that the in-suit

LLMs which emerged from the training process were transformative, OpenAI and Microsoft would have failed to show any reason why they needed to copy Plaintiffs' works for this transformative end. That lack of justification precludes finding that factor one weighs in Defendants' favor.

While Defendants may argue that it was "reasonably necessary" for OpenAI to copy huge volumes of data—and in this motion Plaintiffs must treat that as true—that is not the same as a cognizable justification connecting copying Plaintiffs' works to the LLMs' ostensibly transformative purpose or why the "huge volumes of data" needed to include either Plaintiffs' books or copyrighted books more generally.

Defendants' argument is akin to the *Warhol* magazine publisher arguing that copying the image was "reasonably necessary" because the magazine needed to fill its pages with content about Prince following his death in order to appeal to consumers. *See* 598 U.S. at 532-33. Such copying "often is" "helpful," and such publications may entertain, inform, and enrich the public sphere (and the infringer), but that "does not suffice under the first factor." *Id.* at 547. Instead, because there is no "objective" connection between Defendants' purportedly transformative purpose and the copying of Plaintiffs' works, and because the LLMs threaten wholesale substitution of Plaintiffs' works, factor one weighs decidedly against OpenAI's use.

iii) *OpenAI's Copying Was Commercial.*

Under *Warhol*, even a "transformative" use "must be balanced against the commercial nature of the use," which weighs more strongly where, as here, the use is more likely to substitute for the original. 598 U.S. at 532-33. There is no dispute that Defendants recognized model writing quality as crucial to their commercial success. OpenAI found [REDACTED]

[REDACTED]
SUF ¶¶723-24. OpenAI's "north star" was to get more and more people every week using their

LLMs for writing, a goal they measured using “Writing WAU [weekly active users].” SUP ¶724. Because OpenAI’s gratuitous copying for LLM training had a substitutive purpose and a commercial character, the first factor weighs strongly in Plaintiffs’ favor.

B. Factors Two and Three Favor Plaintiffs: Plaintiffs’ Copyrighted Books are at the Core of Copyright Protection and OpenAI Copied Them In Their Entirety.

As was the case with OpenAI’s piracy, Plaintiffs’ works remain at the core of copyright protection and OpenAI copied the entirety of the works, these factors thus weigh strongly against fair use (particularly given that OpenAI and Microsoft admit they had no justification for copying the entire works in the first place).

C. Factor Four: Defendants’ Use of Plaintiffs’ Works Threatens the Market For Plaintiffs’ Works and Deprives Them of Licensing Revenues.

The fourth fair use factor asks a simple question: What happens to the incentive to write books if conduct like this becomes the norm and is entirely uncompensated? Copyright exists to “provide an economic incentive to create original works,” and the fair use analysis must therefore consider the extent to which a secondary use “diminish[es] th[is] incentive.” *Warhol*, 598 U.S. at 531, 535.

As the proponents of a fair use defense, OpenAI and Microsoft bear the “burden of proving that the secondary use does not compete in the relevant market[s].” *Internet Archive*, 115 F.4th at 189 (citation and quotation omitted). Such evidence must demonstrate the lack of any market effect on “both the market for the original work and the market for any derivative works the rightsholder might develop.” *Id.* In response to such evidence, Plaintiffs need not offer “empirical data” quantifying an impact because courts “routinely rely on . . . logical inferences where appropriate in assessing the fourth fair use factor.” *Id.* at 193; *see id.* (collecting cases).

OpenAI’s exploitation of Plaintiffs’ works to train the GPT models threatens precisely the type of market displacement that is dispositive under the fourth factor. Specifically, if OpenAI’s conduct were widespread, it would significantly harm the market for Plaintiffs’ books (1) by forcing Plaintiffs to compete with AI-generated works for sales, and (2) by eliminating the need to license Plaintiffs’ works in the market for AI training data.

i) Defendants’ Use of Books in AI Training Threatens Wholesale Market Displacement of Human-Authored Books.

The first market harm is the “harm of market dilution.” *Kadrey*, 788 F. Supp. 3d at 1052. Logically, outputs from OpenAI and Microsoft’s GPT-powered products “can still compete for sales with the books in the training data. And by taking sales from those books,” those AI-generated competing works “reduce the incentive for authors to create—the harm that copyright aims to prevent.” *Id.* at 1054. OpenAI used copyrighted works to “train an LLM” that helps to “enable the rapid generation of countless works that compete with the originals, even if those works aren’t themselves infringing.” *Id.* at 1052.

Because Defendants’ conduct poses a fundamental threat to the incentive for human creation, it is firmly within the zone of what factor four is designed to protect. As the Second Circuit recognized in *Google Books*, “copyright is a commercial right, intended to protect the ability of authors to profit from the exclusive right to merchandise their own work.” 804 F.3d at 214. The statute requires the Court to consider the effect of copying on “the potential market for” the copyrighted work and the “value of” it. 17 U.S.C. § 107. As *Kadrey* put it: “[I]ndirect substitution is still substitution: If someone bought a romance novel written by an LLM instead of a romance novel written by a human author, the LLM-generated novel is substituting for the human-written one.” 788 F. Supp. 3d at 1054.

Defendants themselves recognized the threat which their exploitation of copyrighted works without compensation to train the GPT models posed to human authors. OpenAI’s policy director acknowledged that by training the GPT models OpenAI—“an entity of capital with a capital oriented mindset”—was going to “substitute for the labor” of authors; that training the GPT models would make authors “unemployed;” and that the GPT models “*ultimately substitute[] for the practiced, deeply personal labor of the creative communities of the world.*” SUF¶749-50. And Microsoft’s director of applied science noted that continuous exploitation of authors’ work without compensation would result in a “doom-loop” threatening the “economic sustainability of authors” and ultimately resulting in the end of human-authored data for use in training. SUF¶751-68. This “doom-loop” is precisely the harm the Copyright Act seeks to prevent. *See Google Books*, 804 F.3d at 212 (“The ultimate goal of copyright is to expand public knowledge and understanding, which copyright seeks to achieve by giving potential creators exclusive control over copying of their works, thus giving them a financial incentive to create . . . works for public consumption.”).

Bartz’s contrary treatment of market dilution, *see* 787 F. Supp. 3d at 1032, reflects an overly-narrow view of the factor four inquiry not supported by the law or the undisputed facts here. To the extent *Bartz* can be read to suggest that displacement of authors’ works resulting from LLM construction is categorically non-cognizable, that error rests on a flawed premise. LLMs are machines capable of generating an infinite amount of content, with minimal effort and at minimal cost; they are not children learning how to read and write. As Judge Chhabria observed in *Kadrey*, “[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, just because the issue has not come up before.” 788 F. Supp. 3d at 1055. Instead, “[n]o matter how transformative LLM training may be, it’s hard to imagine that it

can be fair use to use copyrighted books to develop a tool to make billions or trillions of dollars while enabling the creation of a potentially endless stream of competing works that could significantly harm the market for those books.” *Id.* at 1059.

The harm to the market for books—precisely the harm OpenAI and Microsoft forecast from OpenAI’s training of the GPT models—is happening now and getting worse. *See* SUF ¶¶706-916. While the parties dispute many facts regarding substitution, it is undisputed: [REDACTED]

[REDACTED] SUF ¶854; that they can be generated in less than ten minutes and for less than a dollar, SUF ¶845; [REDACTED]

[REDACTED] SUF ¶851-853; that these AI books are competing at the highest levels of sales, SUF ¶864-868; and that [REDACTED]

[REDACTED] SUF ¶851.

Against this record, any “short-term public benefits” of OpenAI’s conduct are “outweighed not only by the harm to [Plaintiffs] but also by the long-term detriments society may suffer if [OpenAI’s] infringing use were allowed to continue.” *Internet Archive*, 115 F.4th at 196. The “dearth of creative activity” which would result from smothering the incentive for human authorship in a deluge of AI-produced books would “undoubtedly negatively impact the public.” *Id.* at 195.

ii) OpenAI’s Unlicensed Use Deprived Plaintiffs of Remuneration from an Existing Licensing Market.

Defendants cannot meet their factor four burden for another reason: The deprivation of licensing revenues from the existing, growing market to license works for LLM training. Factor four considers harm to the “traditional, reasonable, or likely to be developed markets” for a given use, *Texaco*, 60 F.3d at 930, including the “impact on potential licensing revenues,” *Internet*

Archive, 115 F.4th at 192. Infringers have no right to “usurp a further market that directly derives from reproduction of the Plaintiffs’ copyrighted works,” *UMG Recordings, Inc. v. MP3.Com, Inc.*, 92 F. Supp. 2d 349, 352 (S.D.N.Y. 2000), particularly where, as here, a licensing market for books for LLM training already exists, *see Texaco*, 60 F.3d at 930-31 (finding harm to licensing revenues for photocopying license weighed against defendant).

The record in this case—and the public record—confirms that the market for licensing works for use in LLM training is robust and growing significantly. Just looking to publicly disclosed agreements, there have already been more than 120 licenses between content providers and AI companies. SUF¶828. Across the industry AI companies are signing licenses to use books in AI training or providing compensation for prior use in training, a collective outlay well-north of \$1 billion. SUF¶828-40. And multiple third-party platforms have emerged to offer content for aggregate AI licensing, including the very same organization—the Copyright Clearance Center—whose photocopying licenses were dispositive of the fourth factor in *Texaco*, 60 F.3d at 930-31. SUF¶831-36.

Both OpenAI and Microsoft have entered into such agreements. [REDACTED]

[REDACTED]

[REDACTED]

[REDACTED] SUF¶814, 821. For its part, Microsoft signed numerous contracts with book publishers, including five agreements in 2024 alone, which provided the right to use books in AI training. SUF¶792. [REDACTED]

[REDACTED]

SUF¶801-07.

The record also confirms that this is a sustainable market, with buyers willing to pay top dollar. OpenAI employees were willing to pay “10X” the billions of dollars they paid for “training a model” to get high-quality books datasets. *SUF*¶297. That was for the simple reason that high-quality books reduced the cost of training substantially and, unlike “compute,” did not depreciate but could be used again and again. *SUF*¶297. [REDACTED]

[REDACTED] *SUF*¶788.

[REDACTED] a. *SUF*¶789. And, as noted above, Microsoft enacted its conviction by signing numerous licenses to access book content for training. *SUF*¶792.

OpenAI’s refusal to pay for training licenses despite the existence of AI training licenses, the avowed utility of books, and their (and Microsoft’s) stated willingness to pay, thus harmed the market for AI training licenses for Plaintiffs’ works and provides an independent reason factor four weighs against OpenAI’s training use.

IV. Defendants’ Distribution of Plaintiffs’ Works to Each Other and Third Parties Was Not Fair Use.

In addition to copies made by downloading from pirate sites and during LLM training, OpenAI and Microsoft also distributed copies of Plaintiffs’ works to each other and third parties. *See supra* Factual Background §§ III.B, IV.C, D. This use violated Plaintiffs’ distribution rights, *see* 17 U.S.C. § 106(3), and neither the distributions to each other nor to third-parties was fair use.

A. Factor One: Defendants’ Transfers Were Substitutive, Not Transformative.

Defendants “create[d] digital copies of the Works and distribute[d] these copies to” each other and third parties “in full, for free.” *Internet Archive*, 115 F.4th at 181. These copies were not transformative. *See id.*

Instead, Defendants used Plaintiffs’ books as currency to facilitate “horse-trading” in their relationship. *SUF*¶589; they sent pirated books to third-parties without maintaining records of what the third-parties did with the works, *SUF*¶607-11; and OpenAI even paid Microsoft ██████████ to download and transfer books as part of Project Mango, *SUF*¶600. There is no question that the materials transferred would, if “wide[ly] disseminat[ed],” pose a significant “risk of substitution” for Plaintiffs’ works. *Warhol*, 598 U.S. at 532.

Nor does some potentially transformative downstream use by recipients justify Defendants’ transfers. *See supra* Argument § II.A. Defendants’ transfers were akin to someone paid to bootleg books for a customer claiming to be a student; the customer’s potential educational use would not justify the bootlegger’s conduct. *See Basic Books*, 758 F. Supp. at 1531.

B. Factors Two & Three Favor Plaintiffs.

As above, these factors weigh in favor of Plaintiffs.

C. Factor Four: Unlicensed Distribution of Plaintiffs’ Works Threatens Complete Destruction of Plaintiffs’ Traditional Markets.

As with OpenAI’s piracy, if distributing unlicensed books for free as part of any commercial relationship or third-party use were widespread, it would destroy the market for Plaintiffs’ works. *See Internet Archive*, 115 F. 4th at 193 (“self-evident that if [distributing and downloading copyrighted books for free] were to become widespread, it would adversely affect [the] markets for the Works in Suit.”); *Mp3Board, Inc.*, 2002 WL 1997918, at *13 (such conduct “on [its] face could harm the market for the original works.”).

All four factors thus weigh decidedly against Defendants’ distribution of Plaintiffs’ works.

V. Microsoft Is Vicariously Liable for OpenAI's Infringement.

In addition, to its direct liability for distributing Plaintiffs' works, Microsoft is vicariously liable for OpenAI's direct infringements.⁴

First, Microsoft had “the right and ability to supervise” OpenAI's infringing conduct, *EMI Christian*, 844 F.3d at 99, because it had control over the Azure servers on which OpenAI downloaded its data, stored its data, trained its models, and deployed its models, and because the agreements it signed with OpenAI gave it the right and ability to prevent OpenAI's infringing conduct. *See supra* Factual Background § IV.A. Microsoft failed to stop or limit OpenAI's infringement despite knowing about them. *See supra id.*; *In re Frontier*, 658 B.R. at 293 (defendant must exercise its control rights to “fullest extent” to avoid vicarious liability).

Second, Microsoft had a “direct financial interest” in OpenAI's infringing actions. OpenAI paid Microsoft for Azure usage and Microsoft benefitted directly from OpenAI's revenues and from the subscribers and users which OpenAI's infringing uses attracted to its Azure and Copilot products. *See supra* Factual Background § IV.A; *EMI Christian*, 844 F.3d at 99 (increase in subscribers/customers due to infringement qualifies as “obvious and direct financial interest.”). Microsoft is thus vicariously liable for OpenAI's infringements.

CONCLUSION

Because OpenAI and Microsoft's conduct strikes at the very core of copyright protection, the Court should grant Plaintiffs' motion for partial summary judgment: (1) that Plaintiffs have established a *prima facie* case of copyright infringement, (2) that OpenAI's downloading of pirated copies from LibGen and via Books3 is not fair use, (3) that OpenAI's further copying of Plaintiffs' works to train its Large Language Models is not fair use, (4) that OpenAI and Microsoft's

⁴ Plaintiffs also have a contributory infringement claim against Microsoft that is unsuitable for resolution at summary judgment as it depends on disputed factual questions.

distribution of copies of Plaintiffs' works to each other and third parties is not fair use, and (5) that Microsoft is vicariously liable for OpenAI's direct infringements.

Dated: September 4, 2026

/s/ Justin A. Nelson

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

CERTIFICATE OF WORD COUNT

In accordance with Local Civil Rule 7.1(c) and the Court's Order at Dkt. 1664, 25-md-3143, I certify that the foregoing Memorandum of Law is 12,450 words, including footnotes but exclusive of the caption page, table of contents, table of authorities, certificates and signature block. The basis of my knowledge is the word count feature of the word-processing system used to prepare this memorandum.

/s/J. Craig Smyser
Signature

CERTIFICATE OF SERVICE

I hereby certify that on September 4, 2026, I caused the foregoing document to be electronically filed with the Clerk of the United States District Court for the Southern District of New York using the CM/ECF system, which shall send electronic notification to all counsel of record.

/s/J. Craig Smyser

(Signature)