

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

IN RE:

OpenAI, INC.,

25-md-3143 (SHS) (OTW)

COPYRIGHT INFRINGEMENT LITIGATION

This Document Relates To:

1:23-cv-11195-SHS-OTW

1:24-cv-01515-SHS-OTW

1:24-cv-03285-SHS-OTW

1:24-cv-04872-SHS-OTW

1:25-cv-04315-SHS-OTW

NEWS PLAINTIFFS' COMBINED SUMMARY JUDGMENT BRIEF

REDACTED VERSION

TABLE OF CONTENTS

I.	Introduction.....	1
II.	Background.....	3
A.	Plaintiffs and Their Works.....	3
B.	Defendants’ Copying of Plaintiffs’ Works for Generative AI.....	5
1.	Data Acquisition	8
2.	Model Training	11
3.	Output Grounding	13
4.	Output Copying.....	15
5.	██████████	31
III.	Defendants directly infringed the asserted works.....	33
A.	Plaintiffs’ Ownership of Valid Copyrights	33
B.	Defendants’ Infringements.....	35
1.	Acquisition, Training, and ██████████	35
i)	The Times and DNP.....	36
ii)	CIR and The Intercept.....	39
iii)	ZD.....	40
2.	Copilot’s RAG Pipeline and Outputs.....	43
i)	The Times and DNP.....	44
ii)	CIR.....	49
3.	Defendants’ Volitional Conduct	50
IV.	Defendants’ fair use defenses fail as a matter of law.....	50
A.	Factor 1: The purposes and character of the uses at issue are substitutive and commercial.	51

1.	The substitutive purposes of defendants’ copying weigh against fair use.....	52
2.	Defendants’ commercial purposes weigh against fair use.	55
3.	The character of Defendants’ conduct weighs further against fair use.....	56
B.	Factor 2: The expressive nature of the asserted works disfavors fair use.	57
C.	Factor 3: The amount and substantiality of the portions used are unreasonable in relation to the purpose of the copying.	57
D.	Factor 4: Defendants cannot show lack of harm to the market for and value of Plaintiffs’ works.....	59
1.	Defendants’ generative AI models and products substitute for news..	65
i)	Defendants’ products output news content.	65
ii)	Defendants’ products do not drive traffic to sources.	67
iii)	Defendants’ products generate pink slime.	69
iv)	Defendants’ products diminish demand for Plaintiffs’ content.....	70
2.	Defendants’ uses supersede feasible licensing markets.....	71
3.	Widespread and unrestricted adoption of Defendants’ conduct would severely harm the market for news along with AI itself.	72
E.	None of the uses at issue in this case are fair.....	74
V.	The Court should grant summary judgment to plaintiffs on the first three elements of their DMCA claims.	75
A.	Plaintiffs conveyed CMI in connection with copies of their Asserted Works.....	76
B.	OpenAI removed Plaintiffs’ CMI without authority.	77
C.	OpenAI’s removal was intentional.	78
VI.	Plaintiffs are entitled to separate awards of statutory damages for infringement of each article.....	78

VII.	The Court should grant summary judgment on Defendants’ remaining affirmative defenses.	80
VIII.	Conclusion	80

TABLE OF AUTHORITIES**Page(s)****Cases**

<i>Abdin v. CBS Broadcasting Inc.</i> , 971 F.3d 57 (2d Cir. 2020).....	41
<i>ADR Int’l Ltd. v. Inst. for Supply Mgmt. Inc.</i> , 667 F. Supp. 3d 411 (S.D. Tex. 2023)	87
<i>American Geophysical Union v. Texaco Inc.</i> , 60 F.3d 913 (2d Cir. 1994).....	<i>passim</i>
<i>Andy Warhol Found. for Visual Arts, Inc. v. Goldsmith</i> , 11 F.4th 26 (2d Cir. 2021)	9, 62, 71, 74
<i>Andy Warhol Found. for Visual Arts, Inc. v. Goldsmith</i> , 598 U.S. 508 (2023).....	9, 61, 71
<i>Associated Press v. Meltwater U.S. Holdings, Inc.</i> , 931 F. Supp. 2d 537 (S.D.N.Y. 2013).....	<i>passim</i>
<i>Authors Guild, Inc. v. HathiTrust</i> , 755 F.3d 87 (2d Cir. 2014).....	44, 61, 62
<i>Authors Guild v. Google, Inc. (“Google Books”)</i> , 804 F.3d 202 (2d Cir. 2015).....	<i>passim</i>
<i>Campbell v. Acuff-Rose Music, Inc.</i> , 510 U.S. 569 (1994).....	<i>passim</i>
<i>Cartoon Network LP, LLLP v. CSC Holdings, Inc.</i> , 536 F.3d 121 (2d Cir. 2008).....	60
<i>Castle Rock Ent., Inc. v. Carol Pub. Grp. Inc.</i> , 150 F.3d 132 (2d Cir. 1998).....	72
<i>Celotex Corp. v. Catrett</i> , 477 U.S. 317 (1986).....	91
<i>Elohim EPF USA, Inc. v. 162 D & Y Corp.</i> , 707 F. Supp. 3d 372 (S.D.N.Y. 2023).....	41, 42
<i>EMI Christian Music Grp.. Inc. v. MP3tunes, LLC</i> , 844 F.3d 79 (2d Cir. 2016).....	90

<i>Fox News Network, LLC v. Tveyes, Inc.</i> , 883 F.3d 169 (2d Cir. 2018).....	61, 73
<i>Gattoni v. Tibi</i> , LLC, 254 F. Supp. 3d 659 (S.D.N.Y. 2017)	86
<i>Google v. Oracle</i> , 593 U.S. 1 (2021).....	9, 66, 75
<i>Grant v. Trump</i> , 749 F. Supp. 3d 423 (S.D.N.Y. 2024).....	42
<i>Hachette Book Group v. Internet Archive</i> , 115 F.4th 163 (2d Cir. 2024)	<i>passim</i>
<i>Harper & Row Publishers, Inc. v. Nation Enters.</i> , 471 U.S. 539	63
<i>Janik v. SMG Media, Inc.</i> , 2018 WL 345111 (S.D.N.Y. Jan. 10, 2018)	87
<i>Kadrey v. Meta</i> , 788 F.Supp.3d 1026 (N.D. Cal. 2025)	<i>passim</i>
<i>Keizai Shimbun, Inc. v. Comline Bus. Data, Inc.</i> , 166 F.3d 65 (2d Cir. 1999).....	52
<i>Knitwaves, Inc. v. Lollytogs Ltd. (Inc.)</i> , 71 F.3d 996 (2d Cir. 1995).....	87
<i>Kregos v. Associated Press</i> , 937 F.2d 700 (2d Cir. 1991).....	42
<i>Mango v. BuzzFeed, Inc.</i> , 970 F.3d 167 (2d Cir. 2020).....	86
<i>MCA, Inc. v. Wilson</i> , 677 F.2d 180 (2d Cir. 1981).....	63
<i>Morris v. Business Concepts, Inc.</i> , 259 F.3d 65 (2d Cir. 2001).....	42, 89
<i>N.Y. Times Co. v. Microsoft Corp.</i> , 777 F. Supp. 3d 283 (S.D.N.Y. 2025).....	86, 89
<i>Nicklen v. Sinclair Broad. Grp., Inc.</i> , 551 F. Supp. 3d 188 (S.D.N.Y. 2021).....	70

<i>NXIVM Corp. v. Ross Inst.</i> , 364 F.3d 471 (2d Cir. 2004).....	63
<i>Pierson v. Infinity Music & Ent., Inc.</i> , 300 F. Supp. 3d 390 (D. Conn. 2018).....	87
<i>Romanova v. Amilus Inc.</i> , 138 F.4th 104 (2d Cir. 2025)	43, 72, 76
<i>Roy Exp. Co. Establishment of Vaduz v. CBS</i> , 503 F. Supp. 1137 (S.D.N.Y. 1980).....	63
<i>Tardif v. City of New York</i> , 991 F.3d 394 (2d. Cir. 2021).....	91
<i>Thomson Reuters Enter. Ctr. GMBH v. Ross Intel. Inc.</i> , 765 F. Supp. 3d 382 (D. Del. 2025).....	65, 72
<i>Twin Peaks Prods., Inc. v. Publications Int’l, Ltd.</i> , 996 F.2d 1366 (2d Cir. 1993).....	66
<i>United States v. Woods</i> , 571 U.S. 31 (2013).....	73
<i>Urbont v. Sony Music Entertainment</i> , 831 F.3d 80 (2d Cir. 2016).....	41
Statutes	
17 U.S.C. § 101	42, 87
17 U.S.C. § 106(1)	43, 51
17 U.S.C. § 107	61
17 U.S.C. § 107(2)	68
17 U.S.C. § 107(3)	68
17 U.S.C. § 107(4)	69, 73
17 U.S.C. § 410(c)	41
17 U.S.C. § 412	90
17 U.S.C. § 504(c)(1).....	89
17 U.S.C. § 1202(b)(1)	11, 86

17 U.S.C. § 1202(c)	87
17 U.S.C. § 1202(c)(1)-(6)	87
17 U.S.C. § 1202(c)(7)	87

Rules

Fed. R. Civ. P. 56(g)	86
Fed. R. Civ. P. 1006	43, 46

Other Authorities

Jane C. Ginsburg, <i>Fair Use Fair Factor Four Revisited: Valuing the “Value of the Copyrighted Work,”</i> 67 J. COPYRIGHT SOC’Y U.S.A. 19 (2020)	73
4 Melville B. Nimmer & David Nimmer, NIMMER ON COPYRIGHT § 13F.08 (2023)	70
United States Copyright Office, COPYRIGHT AND ARTIFICIAL INTELLIGENCE, PART 3: GENERATIVE AI TRAINING (May 2025)	74

I. Introduction

This case is about, as Microsoft’s Director of Applied Science put it, [REDACTED] SF1437, perhaps the [REDACTED] SF1652. Defendants repeatedly copied [REDACTED] of Plaintiffs’ copyrighted articles in their entirety without permission to produce substitutive commercial AI products. OpenAI’s Head of ChatGPT wrote that [REDACTED] from those products, SF1466, which, he said, [REDACTED] and [REDACTED] SF1473-74. Such admissions eviscerate Defendants’ “fair use” defense because substitution is “copyright’s bête noire.” *Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith*, 598 U.S. 508, 528 (2023). For Defendants to prevail on this defense “would,” the same Microsoft executive recognized, [REDACTED] SF1450.

The “goal of copyright” is to “provide an economic incentive to create original works” by granting exclusive rights to authors. *Warhol*, 598 U.S. at 535. Since the Copyright Act of 1976 codified fair use as a narrow limitation on these rights, the Supreme Court has found commercial copying fair only once: for copying code to implement functional compatibility with software interfaces. *Google v. Oracle*, 593 U.S. 1 (2021). The Second Circuit has absolved widespread copying of expressive works only to the extent necessary to allow people to find those works themselves. *Authors Guild v. Google, Inc.* (“*Google Books*”), 804 F.3d 202 (2d Cir. 2015). Defendants’ competitive exploitation of Plaintiffs’ expressive works cannot be excused under these precedents.

To the contrary, in *Kadrey v. Meta*, Judge Chhabria concluded that plaintiffs “will often win” when their works are used for substitutive AI products, with cases involving news publishers “present[ing] even stronger arguments against fair use” than many. 788 F.Supp.3d 1026, 1059 (N.D. Cal. 2025). The record here bears this out with compelling evidence of substitution.

OpenAI's Co-Founder and President, Greg Brockman, wrote that models are [REDACTED]
[REDACTED] SF872,
1425, 1427. Microsoft's CEO Satya Nadella agreed under oath that conversing with chatbots [REDACTED]
[REDACTED]
[REDACTED] SF1432. A Microsoft document recognizes that [REDACTED]
[REDACTED]
[REDACTED]
[REDACTED]
[REDACTED] SF1672.

Defendants' own data show this [REDACTED] in motion, with Microsoft recording [REDACTED]
drops in click-through rates for The Times and DNP's domains, and [REDACTED] for ZD's
domains, for Microsoft's Copilot "answer engine" compared to traditional Bing search. SF1536.
As Microsoft's lead counsel put it, the advantage of such AI chatbots "is that they are actually
designed to answer your question" rather than "just giving you a series of links to things you can
go look for yourself and maybe find the information," SF1511-13—that is, they divert users from
visiting publishers who created the copied work. This is confirmed by multiple third-party reports
documenting the explosive growth in ChatGPT's market share for search, and the low click-
through rates for ChatGPT search results. SF1520-21, 1539.

In light of this record, Plaintiffs seek summary judgment of liability for unauthorized
copying at each stage of the AI pipeline: (1) Acquisition: obtaining copies of Plaintiffs' articles,
often from behind paywalls or in violation of industry norms or terms of use; (2) Training:
configuring models to predict the copied articles' expressive journalistic content; (3) Grounding:
as against Microsoft, using the Bing index to obtain copies to feed to the trained models to generate

real-time responses to user prompts; (4) Outputs: as against Microsoft, reproducing copies or derivatives in model outputs in response to user queries;¹ and (5) [REDACTED] supplying each other with copies for monetary or other consideration as part of Defendants' commercial partnership. There is no dispute that Defendants participated in each type of copying. And they cannot carry the burden of proving their fair use defense for any of them.

In addition, this motion seeks summary judgment of prima facie liability for infringement in Section III, and partial summary judgment that OpenAI intentionally removed Plaintiffs' copyright management information conveyed with their works under the DMCA, 17 U.S.C. § 1202(b)(1), in Section V. Section VI seeks summary judgment that Plaintiffs are entitled to separate statutory damages for each article infringed. And Section VII seeks summary judgment against affirmative defenses for which Defendants have produced no evidence.

Resolution of these issues for Plaintiffs as a matter of law will streamline this case for trial and, by rejecting Defendants' efforts to recast free riding as fair use, recognize that the future not just of journalism but of responsible AI too depends on preserving incentives for humans to produce the creative works on which a healthy society depends.

II. Background

A. Plaintiffs and Their Works

Since our nation's founding, citizens have looked to journalism to discern fact from fiction, understand the news of the day, exercise their rights as citizens, and navigate their lives. Plaintiffs' journalism serves these needs through a wide range of informative works, including coverage of current events, investigative reporting, sports journalism, tech journalism, product reviews and

¹ The motion seeks summary judgment on grounding and output infringement against Microsoft only as the issue is not ripe for determination against OpenAI until the resolution of Plaintiffs' pending sanctions motion.

criticism, service journalism, enthusiast media, and trade reporting. SF10-11, 14, 25, 29, 37-38, 43, 47, 51, 55, 58, 64, 68, 74.

- The New York Times Company (“The Times”) is a global media organization with a staff of 6,000, including over 3,000 newsroom employees, that has produced high-quality news since 1851 and has won 149 Pulitzer Prizes. SF1-12.
- The Daily News Plaintiffs (“DNPs”) are publishers of eight historic newspapers—*New York Daily News*, *Chicago Tribune*, *Orlando Sentinel*, *Sun-Sentinel*, *Denver Post*, *Pioneer Press*, *Orange County Register*, and *Mercury News*, that serve local communities and some larger metropolitan areas with influential and consequential journalism, winning 44 Pulitzer Prizes. SF13-52.
- Ziff Davis (“ZD”) is a diversified digital media company that operates a portfolio of leading vertical media brands, including prominent publications such as IGN, Everyday Health, CNET, ZDNET, PCMag, and Mashable that produce original, professionally written and edited news and analysis, focusing on technology, shopping, gaming, entertainment, health, and wellness. SF53-63.
- The Center for Investigative Reporting, Inc. (“CIR”) is the oldest nonprofit investigative newsroom in the country. Its purpose is to benefit the public through investigative stories about underrepresented voices in our democracy. CIR reaches millions of audience members through its Mother Jones and Reveal brands. SF64-67.
- The Intercept Media, Inc. (“The Intercept”) is a nonprofit news organization dedicated to holding the powerful accountable through fearless, adversarial

journalism. Its investigations focus on politics, war, surveillance, corruption, the environment, technology, criminal justice, the media, and other issues. SF68.

B. Defendants’ Copying of Plaintiffs’ Works for Generative AI

OpenAI formed in December 2015 as a nonprofit purporting to “benefit humanity as a whole.” SF627-29. Yet within two years, its founders were contemplating commercialization in earnest. No later than 2017, co-founder and CEO Sam Altman [REDACTED] [REDACTED] and OpenAI [REDACTED] [REDACTED] SF633-34; *see* SF635-37. Around the same time, OpenAI co-founder Greg Brockman wrote he was [REDACTED] [REDACTED] SF630. Lately, it has been reported that OpenAI is planning an IPO based on a \$1 trillion valuation. SF654.

In pursuing its for-profit mission, OpenAI has copied [REDACTED] of Plaintiffs’ articles to build many large language models, including GPT-3, GPT-3.5, GPT-4, GPT-3.5 Turbo, GPT-4 Turbo, GPT-4o, and GPT-4o mini. SF563-64, 838, 848-49; *see generally* SF828-919. These models underlie OpenAI’s commercial products, including the OpenAI API (GPT-3) released in 2020, and its blockbuster ChatGPT chatbot (GPT 3.5) released in 2022. SF552-553, 562, 567. By February 2026, there were [REDACTED] weekly ChatGPT users, many of whom use ChatGPT to access or receive news content. SF575, 1479-1489, 1522.

Another commercial OpenAI product is its “CustomGPT” platform and store, which hosts projects like the “News Summarizer Ace,” “NYTimesGPT,” and “Bypass Paywall,” pictured below (red emphasis added). SF588-596, 1015.



News Summarizer Ace

By MR MW KAHN &

Articles, Videos and Podcasts: Concise briefs from any URL, **paywall or not**, with our GPT! Select word length and language preference effortlessly.

Productivity

Category

300+

Conversations

Conversation Starters

Skip paywalls by just using the link text or URL.

Bypass tedious copy-past or video watch or listen.

Save on subscription costs.

Our GPT guides your reading, watching, listening choices.

Bypass Paywall (by Paywall.vip)

By Niklas D. Palladini &

Access paywalled articles easily. Share your URL and get a Paywall.vip bypass link instantly.

Share an article
URL - I'll give
you the bypass



Other Custom GPTs include projects named “Remove Paywall”; “Article Reader” (“Use this GPT to summarize and extract key points from paid (paywalled items). Just submit the URL”); and “Article and News Analyzer GPTlee 1.0” (“A Tool for Extracting stories and data from your daily news articles and getting no ads—just add the hyperlink or article title”). SF1470.

Since 2019, Microsoft has partnered with OpenAI to [REDACTED] SF638-639. In exchange for a [REDACTED] share of revenue worth [REDACTED] and equity worth [REDACTED], Microsoft has provided [REDACTED] in funding. SF654, 661-664, 756, 765. Microsoft has also provided extensive computing infrastructure, [REDACTED] and other know how. SF669-742. Sam Altman has admitted that OpenAI was [REDACTED] SF670.

Microsoft retained control over OpenAI’s access to these critical computing resources, including [REDACTED]. SF688, 778. At his deposition, Microsoft’s CEO Mr. Nadella confirmed that Microsoft also had the right and ability [REDACTED] SF768-769. [REDACTED]

[REDACTED] SF771-772.

Instead, in addition to [REDACTED], SF759-760, Microsoft deployed the GPT models within its own commercial products, including its “Bing Chat” chatbot, later rebranded as “Copilot,” SF597-601. Indeed, Microsoft has since deployed OpenAI models into over 500 Microsoft products, including the “Copilot Search” chatbot, the “Recipe Tips” product, the “Bing Shopping” product, and the “Summarize Top News from SharePoint in Viva Connections” product. SF621-624.

1. Data Acquisition

Defendants scraped, downloaded, and otherwise obtained copies of Plaintiffs’ works to build repositories and datasets for many uses.

OpenAI’s expert explained: [REDACTED]

[REDACTED] SF782.

Given the scale of copying at issue here, Defendants used computer programs known as web crawlers or scrapers, or simply “bots” (short for “robots”), to access Plaintiffs’ webpages and extract their contents. SF714, 782-789, 799, 809-811. OpenAI did not [REDACTED]

[REDACTED]. SF521-47, 791-92. For its part, Microsoft [REDACTED]

[REDACTED] SF689–711.

Beginning with the GPT-2 model, OpenAI “created” a new web scrape called WebText, which “emphasizes document quality by prioritizing web pages curated/filtered by humans.” SF801–03. [REDACTED]

[REDACTED] SF804. For the GPT-3 and GPT-3.5 models, OpenAI created an

expanded version of WebText called WebText2, which also [REDACTED] [REDACTED] SF805, 808. Webtext2 includes at least [REDACTED] scraped works from The Times; [REDACTED] from DNP, and [REDACTED] from ZD. SF1184–85, 1223; *see also infra* Section III.B.1.i.

For GPT-3 onwards, OpenAI also downloaded datasets from Common Crawl—an entity that [REDACTED]—containing [REDACTED] of complete copies of Plaintiffs’ website content. SF812-13, 823-824, 1184–85, 1190-91, 1203-12, 1223; *see also infra* Section III.B.1.i. Plaintiffs did not authorize Common Crawl to scrape or disseminate their content. SF1285-99. Even so, Common Crawl’s own terms of use require users to adhere to any underlying terms of use applicable to the data that it scrapes. SF815.

Individuals within OpenAI and Microsoft ignored such issues [REDACTED] [REDACTED]. SF521-47, 790-92. For example, OpenAI’s corporate representative testified that he was [REDACTED] [REDACTED] SF792. While Microsoft’s internal documents acknowledge the [REDACTED] noting that ChatGPT [REDACTED] [REDACTED] [REDACTED] SF626 (emphasis added). Mr. Nadella testified [REDACTED] [REDACTED] SF793. And he agreed that AI companies [REDACTED] SF795.

But that is exactly what Defendants did. SF521–47, 791-92. OpenAI [REDACTED] [REDACTED] [REDACTED] SF813, 822. An OpenAI Slack discussion included a list

of [REDACTED]
[REDACTED]
[REDACTED]. SF844. By acquiring content through third parties such as Common Crawl and through anonymous scrapes, OpenAI also sidestepped the automated robots.txt regime that web publishers use to restrict access to content, acquiring publishers' content at scale in defiance of any efforts to block it from scraping and without regard to any applicable terms of use. SF493, 501, 505.²

As discussed in Section II.B.1 below, Defendants exchanged scraped copies of Plaintiffs' content from the [REDACTED] SF689–740. OpenAI delivered [REDACTED]
[REDACTED]. SF748-50. Microsoft similarly provided training data to OpenAI through initiatives called [REDACTED] and [REDACTED]. SF689–740.
[REDACTED]
[REDACTED]. SF730-31.

OpenAI also obtained from a third-party, Linguistic Data Consortium, a dataset called The New York Times Annotated Corpus (“Annotated Corpus”), which had “nearly every article published in The New York Times between January 1, 1987 and June 19, 2007”—over 1.8 million articles. SF1093. The Annotated Corpus was subject to a User License Agreement that limited use to “non-commercial linguistic education, research and technology development,” and proscribed

² In 2024, OpenAI announced it planned to release a product called “Media Manager” to answer the “need” for “an efficient, scalable solution for content owners to express their preferences about the use of their content in AI systems,” including by allowing publishers to opt out of its scraping. SF1434. It has since [REDACTED]. SF1435.

commercial use. SF1105. OpenAI employees knew [REDACTED]

[REDACTED] SF1107, [REDACTED] SF1116-23.

OpenAI saved copies of Plaintiffs' content in various repositories, including on [REDACTED]

[REDACTED] SF1392. Defendants made additional copies and then used them in several distinct ways, including training, quality filtering, and model validation. SF564, 1116-23, 1204-12. As Microsoft recognized: [REDACTED]

and admitted [REDACTED]

[REDACTED] SF1437.

2. Model Training

Defendants made and used additional copies of Plaintiffs' content for pre-training, mid-training, and post-training the GPT models, as well as for data quality filtering, model evaluation, and building a "Bloom" filter.

Pre-training: The accused LLMs contain billions of parameters (weights) used to generate outputs (responses) to inputs (queries or prompts) one token (i.e., a word or word part) at a time. SF847. These LLMs generate outputs by predicting the next word that is likely to follow an input text based on the similarity of the input to text used to train the model. SF1040. During pre-training, OpenAI passed [REDACTED] of copies of Plaintiffs' articles through the model. SF849. As a result, the models have encoded specific text from those articles in their parameters (a process called "memorization") that the models can output verbatim or near-verbatim ("regurgitation"). SF925-26. OpenAI's training sets [REDACTED]

[REDACTED] SF1136-138, 1300-04, 1363-1421.

Mid-Training: OpenAI used [REDACTED]

[REDACTED]
[REDACTED]. SF876-78, 884. OpenAI's goal was to [REDACTED]
[REDACTED]
[REDACTED] SF886-88. OpenAI
sought to improve [REDACTED]
[REDACTED] SF885, 888. OpenAI's mid-training
datasets contain over [REDACTED] copies of Plaintiffs' works. SF1188-89, 1202, 1205, 1221.

As a result of pre-training and mid-training, the models memorized Plaintiffs' articles.
OpenAI's VP of Research explained: [REDACTED]
[REDACTED] SF927. By November 2019, OpenAI internally [REDACTED]
[REDACTED] SF942. By 2020, OpenAI understood that
[REDACTED]
[REDACTED] SF943. That same year, OpenAI recognized that its [REDACTED]
[REDACTED] SF945. By 2021, OpenAI considered [REDACTED]
[REDACTED] SF946.
In June 2022, OpenAI employees acknowledged that GPT-4 would [REDACTED]
[REDACTED] SF949.

Post-Training: During post-training, Defendants used [REDACTED]

[REDACTED] SF894-97, 920-24.
Defendants' post-training methods [REDACTED]
[REDACTED]
[REDACTED] SF918-19. [REDACTED]

[REDACTED]

SF916-19.

Defendants also used Plaintiffs' content [REDACTED]

[REDACTED] SF923, 1117-23. Additionally, immediately after the initiation of these lawsuits, OpenAI [REDACTED]

[REDACTED] SF1147-48. [REDACTED]

[REDACTED] SF1151. [REDACTED]

[REDACTED]

[REDACTED]

Microsoft's Director of Applied Science, Dr. Brent Hecht, [REDACTED]

[REDACTED]

[REDACTED]

SF1154.

3. Output Grounding

In response to user queries, ChatGPT and Copilot use a technique called retrieval augmented generation ("RAG") to fetch copies of Plaintiffs' content from their websites [REDACTED] to use as sources for generating responses to user queries. SF971-1035, 1229-1256. As Microsoft describes the process: "A basic [RAG] model begins with a trigger, such as a user query or instruction. This trigger is sent to a retrieval function, which fetches relevant content based on the query. The retrieved content is then merged back into the context window of the LLM, along with the input prompt and the query itself. . . Finally, the LLM generates an output based on the combined input and retrieved content." SF972.

³ See https://simple.wikipedia.org/wiki/Bloom_filter.

At least [REDACTED] Copilot conversations copied and grounded model outputs on Plaintiffs' website content as part of the RAG process. SF1234, 1252. Those conversations include unique URLs corresponding to Plaintiffs' Asserted Works: [REDACTED] for The Times and [REDACTED] for DNP. SF1237. Microsoft's technical expert Dr. John Lafferty concedes the point [REDACTED]

[REDACTED] SF1034. Dr. Lafferty further admits that there are at least [REDACTED] instances in which Microsoft copied at least six non-overlapping [REDACTED]s [REDACTED] from one of Plaintiffs' works as part of the RAG process. SF1236.

OpenAI created blocklists to prevent ChatGPT from grounding on certain websites, but did not block any Plaintiff website until [REDACTED], and blocked others as late [REDACTED] SF1019. Despite those blocklists, and even though Plaintiffs followed the instructions that Defendants claimed would let them "opt-out," both [REDACTED] [REDACTED] SF483-505, 507 (OpenAI) and 512-520 (Microsoft).

Defendants marketed their chatbots as alternatives to traditional search engines. SF 1524, 1526. But unlike traditional search engines, Defendants' products: (1) reduce the incentive for users to click on links to publishers' websites, (2) generate outputs using up to [REDACTED] characters of grounded content, and (3) provide portions of paywalled content to users. SF1536 (Microsoft data showing the overall click-through rate difference between Copilot and traditional Bing Search was [REDACTED] for The Times, [REDACTED] for DNP, and [REDACTED] for ZD); SF1524 (Copilot's home page welcoming users by stating: "Instead of clicking through links, we can talk through whatever you're curious about."); SF1035 (Microsoft grounding on at least [REDACTED] words of a single Plaintiff article); SF1025 (OpenAI grounding on an entire Plaintiff article); SF979 (OpenAI researchers explaining in March 2023 that ChatGPT was able to [REDACTED])

[REDACTED]

Verbatim copying from the grounding content systematically occurs in Defendants' chatbot outputs to users, discussed below. SF1010-1012. As Microsoft's expert put it: [REDACTED]

[REDACTED] SF1013.

4. Output Copying

Defendants' chatbots output Plaintiffs' content because they were trained and grounded on that content. OpenAI's technical expert, Dr. Chris Callison-Burch, conceded [REDACTED]

[REDACTED]. SF852. He explained that

[REDACTED]. SF850-851. While the Defendants may dispute the exact numbers, records of actual interactions ("conversations") between the models and Defendants' customers show that [REDACTED]

OpenAI's additional technical expert, Dr. Taylor Berg-Kirkpatrick, found [REDACTED] of verbatim 25-gram spans of text ("25-grams") contained in Plaintiffs' works throughout a tranche of 20 million ChatGPT conversations that OpenAI produced:⁴

	25-grams	Unique ChatGPT Conversations	Unique Plaintiff Articles
NYT	[REDACTED]	[REDACTED]	[REDACTED]
DNP	[REDACTED]	[REDACTED]	[REDACTED]

⁴ While OpenAI disputes whether some of the reproduced content is original and owned by Plaintiffs, its presence in the ChatGPT conversations at least demonstrates the LLMs' propensity to memorize and regurgitate training data. *See* Section II.B.2, *supra* (discussing memorization and regurgitation).

	25-grams	Unique ChatGPT Conversations	Unique Plaintiff Articles
CIR			
The Intercept			
ZD			

SF1059.

In the tranche of 8.2 million Copilot conversations that Microsoft produced, Microsoft's expert Dr. Lafferty likewise found that at least [REDACTED] Copilot outputs each shared a minimum of six non-overlapping 16-grams (96 words) with a Plaintiff work. SF1246.

Users turn to Defendants' products for news, including Plaintiffs' content. OpenAI hosted hundreds of Custom GPTs designed to access, store, provide, or view news content, including Plaintiffs' content protected by paywalls or other access controls. SF590-596, 1468-1470; Section II.B, *supra*. Microsoft encourages users to use Copilot to "[s]ummarize the top news." SF1514.

Defendants' experts asserted that at least [REDACTED] of roughly [REDACTED] user requests to ChatGPT ask for [REDACTED]

[REDACTED] and at least [REDACTED] of roughly [REDACTED] Copilot conversations are directed to [REDACTED]

[REDACTED] SF1481, 1488-1489. OpenAI's expert asserted that another study of July 2025 ChatGPT conversations indicates [REDACTED]

[REDACTED] which when applied to ChatGPT's [REDACTED] prompts a day, is approximately [REDACTED] prompts a day [REDACTED] SF1160. OpenAI's internal data as of February 2026 tracked that [REDACTED]

[REDACTED] SF1669. While these figures [REDACTED]

[REDACTED] SF1478-1500.

The Copilot and ChatGPT conversations Defendants produced also contain substitutive and harmful non-verbatim outputs that result from Defendants having trained and/or grounded on copies of Plaintiffs' content. These outputs include summaries of Plaintiffs' articles, content generated in the style of Plaintiffs' publications, and rewritten Plaintiffs' articles. SF1075, 1078-1079, 1171, 1069-1075, 1078-1092. There is also "hallucinated" content that plausibly but falsely purports to originate from a Plaintiff. SF1075, 1171, 1740-1741.

The Times: By way of example only, in response to the prompt [REDACTED]
[REDACTED] ChatGPT retrieved and output the text of The Times's article. SF1066. The screenshot below highlights The Times's content in red.

⁵ Many of the ChatGPT conversation records produced to Plaintiffs replaced the URL from the actual conversation with "<URL>." SF1076.



In response to the prompt [REDACTED] ChatGPT
retrieved and output the text of another Times article (highlighted in red). SF1073.



In response to a prompt to [REDACTED], ChatGPT invoked the WebPilot Custom GPT—a GPT that OpenAI identified as [REDACTED] — to retrieve the text of the article and then output the verbatim text to the user. SF1015-1018, 1470. The Times’s article text is highlighted in yellow below.





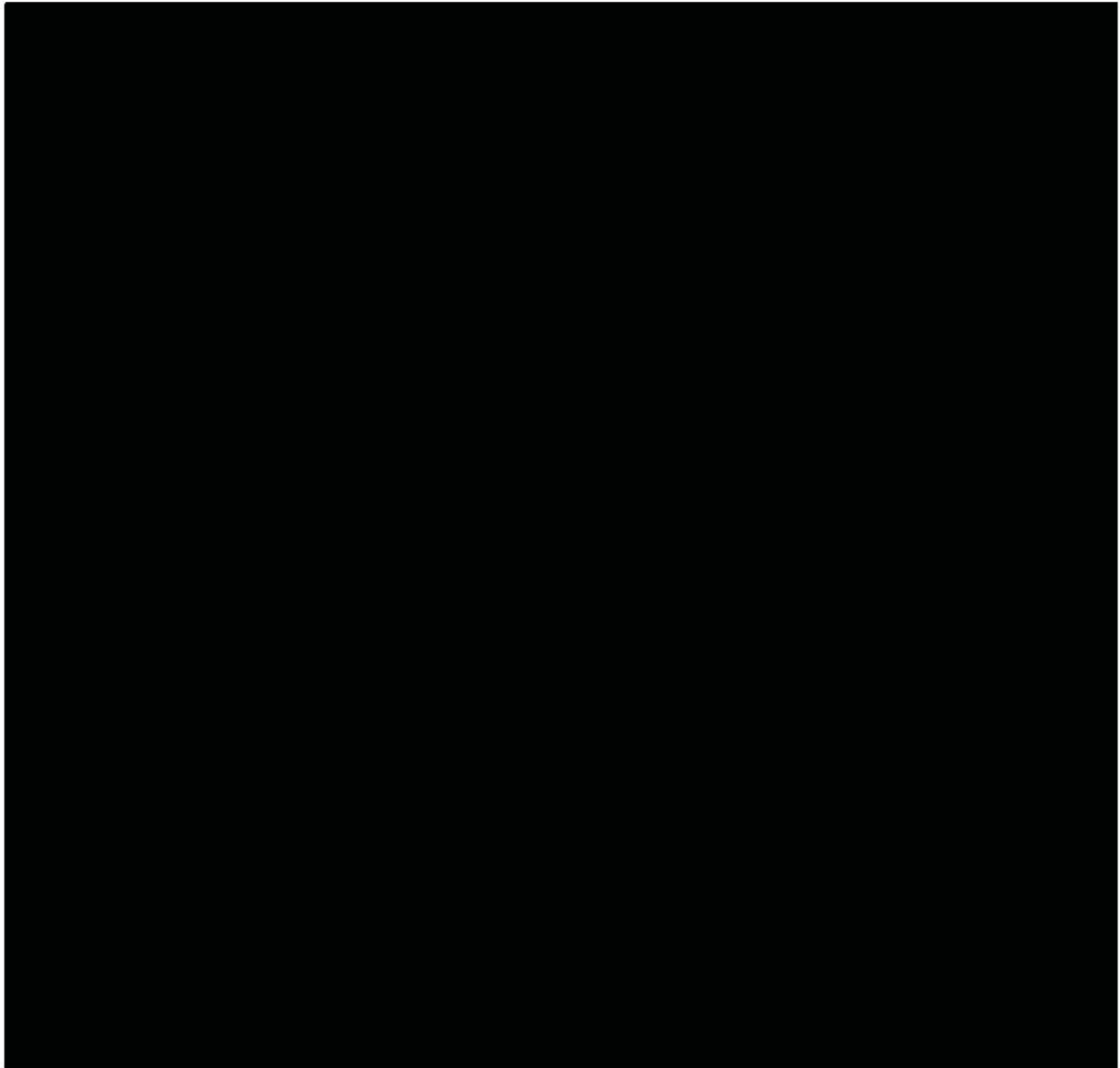
In response to the prompt

[REDACTED]

[REDACTED]

ChatGPT output a seven-paragraph summary of

The Times's article. SF1083.



In response to the prompt:

ChatGPT output a four-paragraph summary of The Times's article. SF1080.



In response to the prompt

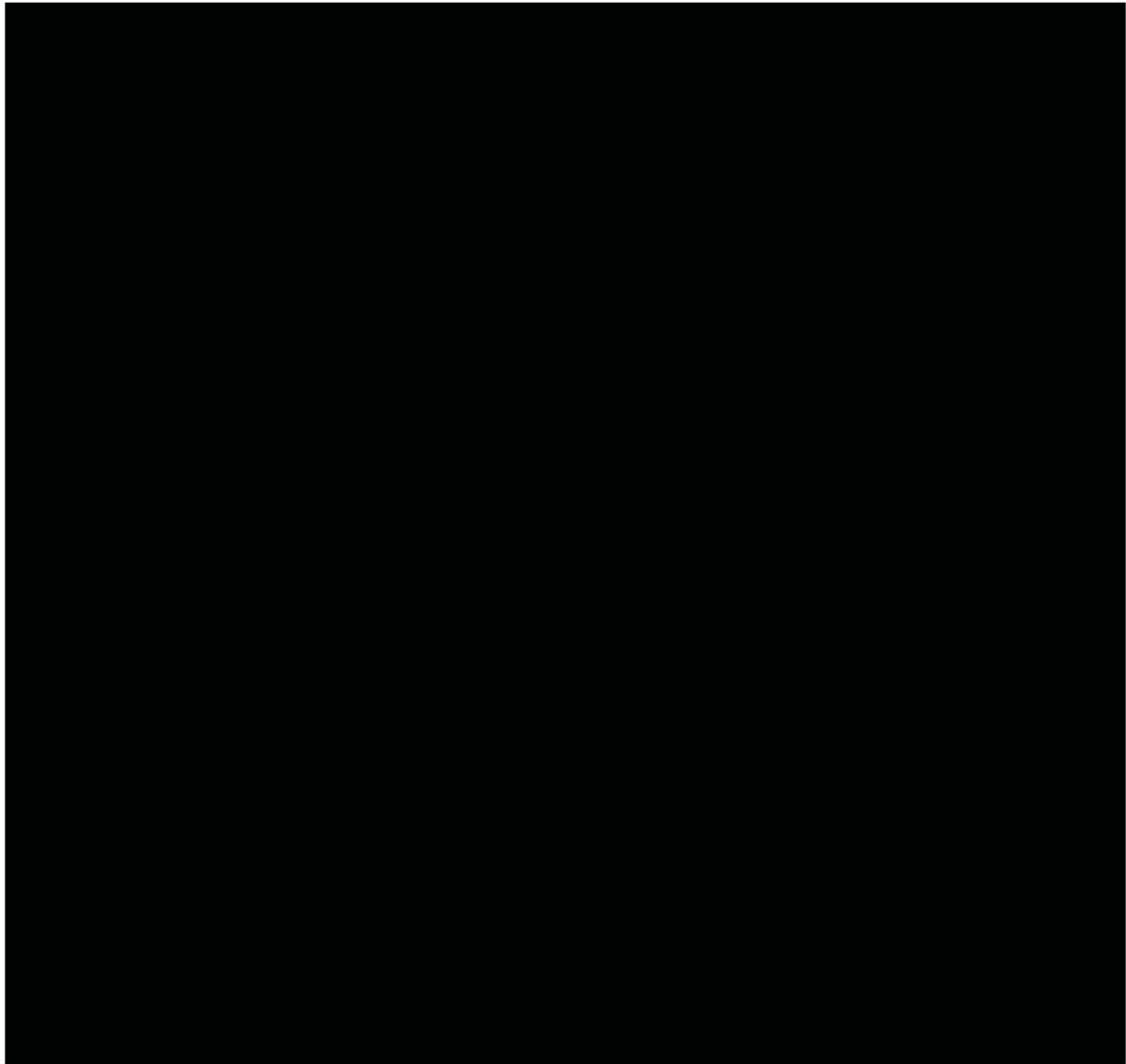


ChatGPT output a four-paragraph

summary of The Times's article. SF1081.



In response to the prompt [REDACTED] ChatGPT output a four-paragraph summary of a Times article. SF1084.



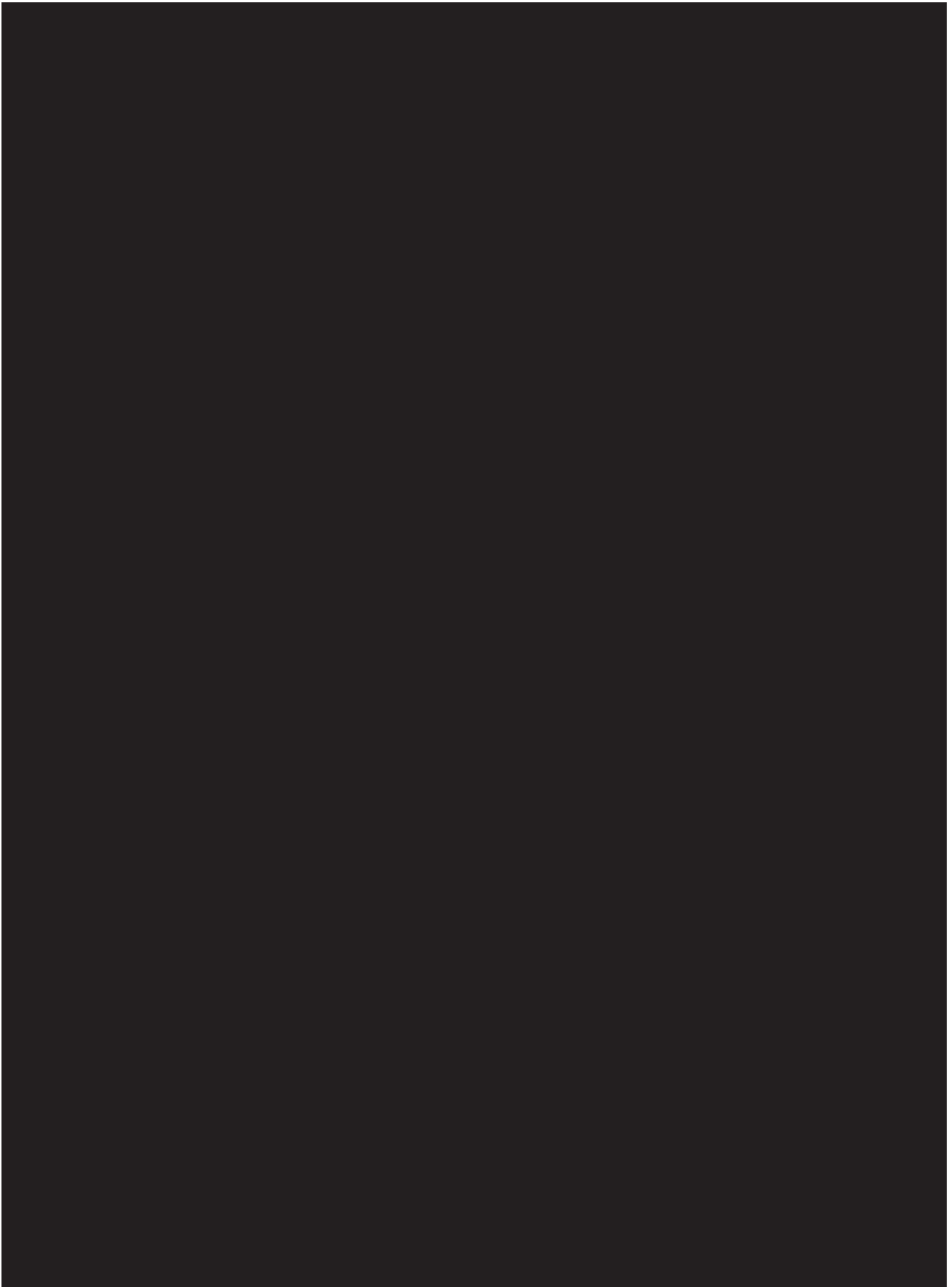
DNP: In response to a prompt [REDACTED]

[REDACTED]

[REDACTED] and a follow up prompt [REDACTED]

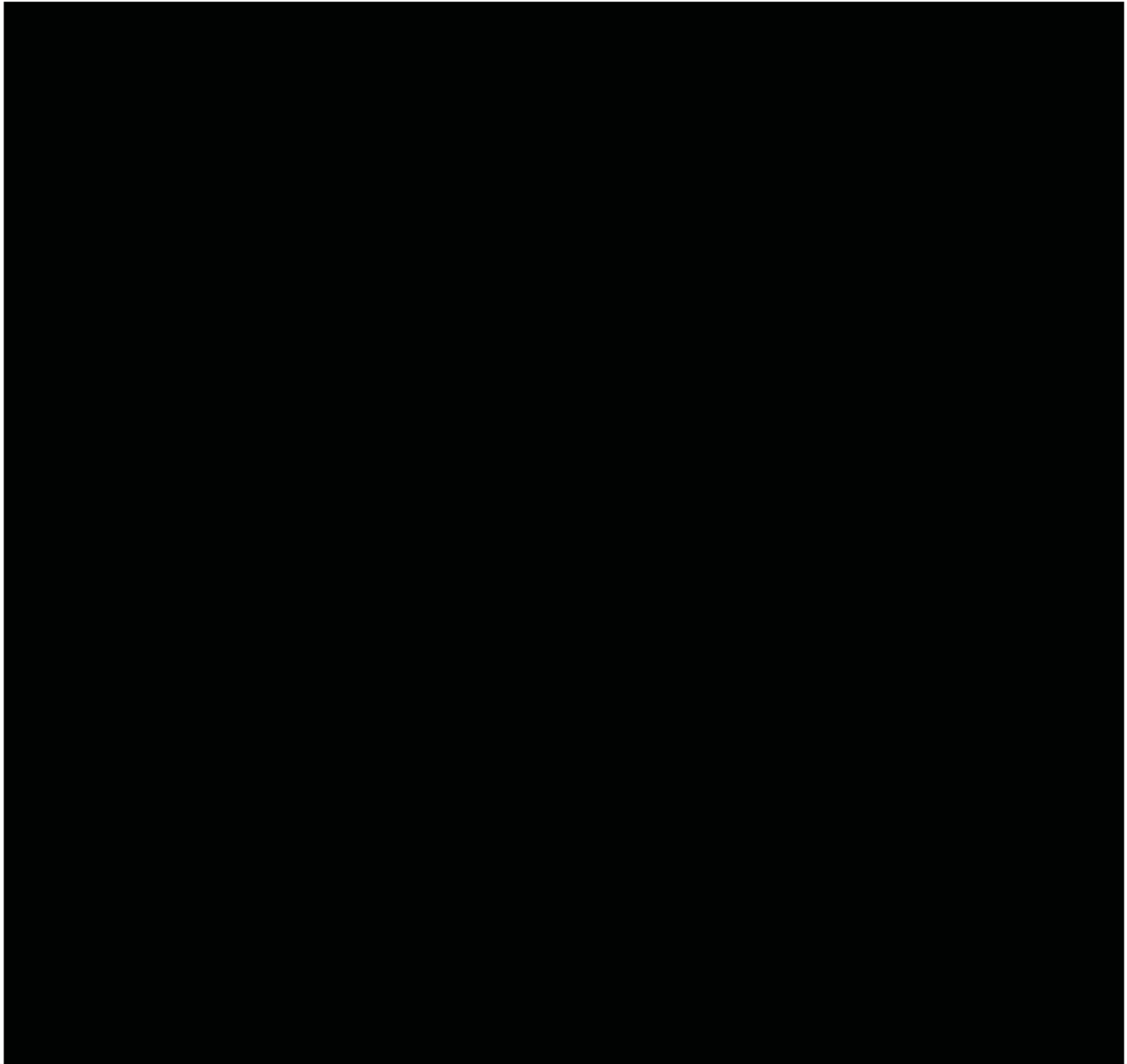
[REDACTED] ChatGPT retrieved and output the text of two Mercury News's articles. SF1067-68.





In response to the prompt [REDACTED]

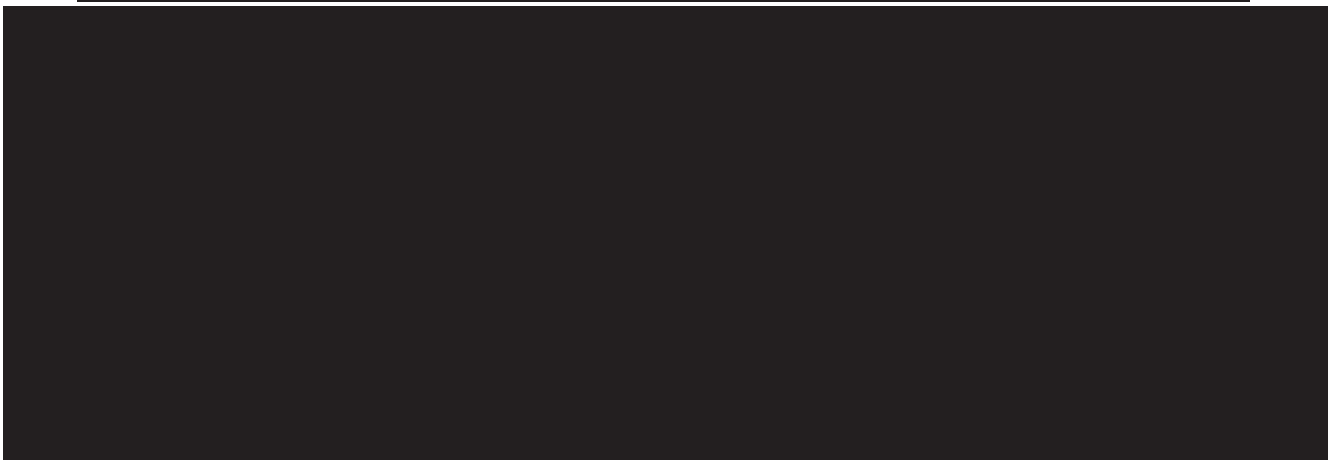
[REDACTED] ChatGPT output a six-paragraph summary of a Mercury News article. SF1082.



In response to the prompt [REDACTED] ChatGPT output a four-paragraph summary of a Denver Post article. SF1084.



ZD: OpenAI's expert (Dr. Berg-Kirkpatrick), showed that in response to various prompts, OpenAI's models outputted verbatim ZD content, such as portions of Lifehacker and Eurogamer articles:



SF1060-1061.

These outputs are not coincidental. They are the consequence of Defendants' copying of Plaintiffs' works during training and grounding through RAG. SF851-852, 872-873, 890, 925-927, 929-932, 937-939, 943, 989, 998, 1018, 1034, 1066-1074. According to OpenAI's expert Dr. Callison-Burch, memorization [REDACTED]

[REDACTED] SF1822 (Dr. Callison-Burch: [REDACTED]

a [REDACTED]

OpenAI did not merely permit its models to summarize Plaintiffs' content; it [REDACTED] SF910-916. Under its [REDACTED] . SF910-916. OpenAI's own testing shows the same behavior: in Dr. Berg-Kirkpatrick's [REDACTED] experiments, ChatGPT [REDACTED]

[REDACTED] . SF1824

Similarly, summarizing news content is one Copilot capability Microsoft advertises to users. SF1514. Microsoft's own expert, Dr. Lafferty, acknowledges that Copilot's outputs [REDACTED] SF1825.

5. [REDACTED]

Defendants supplied copies of publisher content to each other, using Plaintiffs' content as [REDACTED] . SF697-702. Microsoft supplied OpenAI with copies of billions of webpages for use in developing the GPT models, in exchange for [REDACTED] . SF691-706. Instead of paying Plaintiffs to license their content, Defendants [REDACTED]

OpenAI gave [REDACTED] to Microsoft: In September 2020, OpenAI agreed to provide Microsoft the [REDACTED], so that

Microsoft could [REDACTED] SF564, 748. [REDACTED]
 [REDACTED], SF551-552, so this exchange bore no relationship to training any model at issue.
 Microsoft accessed the [REDACTED]
 [REDACTED] SF750. Put
 differently, Microsoft [REDACTED]
 [REDACTED]

Microsoft supplied Bing Index data to OpenAI (Project [REDACTED]): Over a three-year period from 2019 to 2022, Microsoft provided OpenAI with a copy of the “Bing Index,” which Microsoft’s counsel has described as a compilation of “billions” of webpages which Microsoft gathered to support its traditional Bing search engine. SF691-694, 700. This data transfer, which included Plaintiffs’ content, was codenamed [REDACTED] SF695. OpenAI convinced Microsoft to [REDACTED] it the Bing Index for [REDACTED] SF694; *see* SF705. OpenAI thereafter agreed to [REDACTED]
 [REDACTED] SF698, 703. In October 2021, OpenAI’s Chief Research Officer Bob McGrew described [REDACTED]
 [REDACTED] SF699. But Defendants never asked copyright holders for permission [REDACTED]. SF707. And once again, OpenAI had by then already [REDACTED]
 [REDACTED]

Defendants [REDACTED] (“Project [REDACTED]” After providing OpenAI with the Bing Index, Microsoft launched “Project [REDACTED] to help [REDACTED]
 [REDACTED] SF712. Microsoft developed and operated the [REDACTED]

[REDACTED]. SF715-725. [REDACTED]
 [REDACTED] SF729.
 In exchange, OpenAI [REDACTED]. SF720. Yet Microsoft never sought or
 obtained permission from the publishers [REDACTED]
 [REDACTED] SF726.

III. Defendants directly infringed the asserted works.

Plaintiffs move for judgment that they own valid copyrights in their Asserted Works and that: (1) Defendants infringed Plaintiffs' Asserted Works by acquiring, training on, and [REDACTED] those works; and (2) Microsoft infringed Plaintiffs' Asserted Works used as RAG inputs or reproduced in RAG outputs from the consumer Copilot chatbot.

To prove infringement, a copyright plaintiff must show “(1) ownership of a valid copyright, and (2) copying of constituent elements of the work that are original.” *Abdin v. CBS Broadcasting Inc.*, 971 F.3d 57, 66 (2d Cir. 2020) (quoting *Feist Publ'ns, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340, 361 (1991)). “To satisfy the second element, a plaintiff must demonstrate that: (1) the defendant has actually copied the plaintiff's work; and (2) the copying is illegal because a substantial similarity exists between the defendant's work and the protectible elements of plaintiff's work.” *Id.* (cleaned up).

A. Plaintiffs' Ownership of Valid Copyrights

A certificate of registration “made before or within five years after first publication of the work shall constitute prima facie evidence of the validity of the copyright[.]” 17 U.S.C. § 410(c). Once the copyright owner presents a certificate of registration, “[t]he party challenging the validity of the copyright has the burden to prove the contrary.” *Elohim EPF USA, Inc. v. 162 D & Y Corp.*, 707 F. Supp. 3d 372, 392 (S.D.N.Y. 2023) (quoting *Hamil Am. Inc. v. GFI*, 193 F.3d 92, 98 (2d Cir. 1999)). This presumption extends to copyright ownership. *See Urbont v. Sony*

Music Entertainment, 831 F.3d 80, 89 (2d Cir. 2016). “To rebut the presumption of validity, defendants must prove by a preponderance of the evidence that the copyright was not properly registered.” *Elohim*, 707 F. Supp. 3d at 392.

Because a newspaper or magazine is a “collective work,” 17 U.S.C. § 101, those copyright registrations extend to each article within them for which the registrant also owns the copyright. *Morris v. Business Concepts, Inc.*, 259 F.3d 65, 68 (2d Cir. 2001), *abrogated on other grounds by Reed Elsevier, Inc. v. Muchnick*, 559 U.S. 154, 155 (2010); *Grant v. Trump*, 749 F. Supp. 3d 423, 431-32 (S.D.N.Y. 2024); *see also* U.S. Copyright Office, Circ. 62A, “Group Registration of Newspapers,” (rev. Mar. 2021) at 2 (“[B]ecause each issue is registered as a collective work, the registration also covers the articles....”).

For digital media publishers like ZD that operate websites that provide news and other content, the publisher can apply to register individual articles, submit a “GRNW” (for certain websites), or submit a database registration for its editorial databases. Compendium of U.S. Copyright Office Practices (3d Ed.2021) (“Compendium”) §204.3(A), §§1112 et seq.; <https://www.copyright.gov/eco/grnw/help-tx.html>. Database registrations extend to the compilation and to each unpublished work contained in the database. Compendium §204.3(A), 1112 et seq.

There is no material dispute of fact that each of the registered Asserted Works was registered validly under one of these approaches. SF295-305, 316-319, 328-351, 385-388, 395-402; 408-416. Nor is there any dispute each “was independently created by the author... and that it possesses at least some minimal degree of creativity,” as required for copyrightability. *Kregos v. Associated Press*, 937 F.2d 700, 703, 705 (2d Cir. 1991) (quoting *Feist Publ’ns*, 499 U.S. at 345). Nor is there any dispute that each Asserted Work was authored either by an employee or

pursuant to a “work for hire” style agreement to the benefit of the publishing Plaintiff or its predecessor in interest. SF305-315, 317-327, 352-384, 387-394, 424, 431. Rule 1006 summaries set forth the registration and work-for-hire basis for each Asserted Work. SF305, 317-318, 387-388, 410-412, 417-418, 428.

For unregistered works subject to DMCA claims, there is also no dispute about independent creation, creativity, work-for-hire, and ownership. SF408, 410, 417-425, 427-432.

Plaintiffs are entitled to summary judgment on the validity of their copyrights.

B. Defendants’ Infringements

To establish unauthorized copying, “a plaintiff must show both that his work was ‘actually copied’ and that the portion copied amounts to an ‘improper or unlawful appropriation.’” *Jorgensen*, 351 F.3d at 51. Direct evidence of copying is rare, so copying may be inferred when (1) a defendant had access to the original work; and (2) the defendant’s work bears a ‘substantial similarity’ to the original.

1. Acquisition, Training, and [REDACTED]

Defendants infringed Plaintiffs’ exclusive right of reproduction by scraping and downloading the Asserted Works. 17 U.S.C. § 106(1); Section II.B.1, *supra*. The additional copies of the Asserted Works generated during Defendants’ training of the LLMs (Section II.B.2, *supra*), also independently infringed Plaintiffs’ exclusive right of reproduction. *Id.* And the additional copies of the Asserted Works generated during Defendants’ [REDACTED] of those works with each other (Section II.B.5, *supra*) also independently infringed Plaintiffs’ exclusive rights of reproduction and distribution. *Id.* § 106(1), (3).

Where direct evidence shows wholesale copying or nontrivial verbatim overlap of expressive content, courts find infringement and proceed directly to defendants’ defenses. *See, e.g., Romanova v. Amilus Inc.*, 138 F.4th 104, 108–11, 115–22 (2d Cir. 2025) (proceeding to fair

use where the defendant republished plaintiff's photograph); *American Geophysical Union v. Texaco Inc.*, 60 F.3d 913, 915–17, 922–32 (2d Cir. 1994) (analyzing fair use after Texaco scientists made complete photocopies of copyrighted journal articles); *Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87, 91–92, 94–101 (2d Cir. 2014) (addressing fair use of complete digital book copies); *Authors Guild v. Google, Inc.*, 804 F.3d 202, 207–10, 212–25 (2d Cir. 2015) (same). Here, Plaintiffs seek judgment only with respect to the copies of the Asserted Works that Defendants downloaded and further copied into in the training datasets that meet or exceed thresholds that demonstrate extensive verbatim overlap. Plaintiffs will submit any questions of substantial similarity for works that do not meet these thresholds to the jury at trial.

i) The Times and DNP

The Times and DNP seek summary judgment only as to the Asserted Works meeting or exceeding a “coverage score” in which 80% of the 3-grams in the Asserted Work are contained in Defendants’ copy *and* the Asserted Work and the copy share a 16-gram. The Times and DNP will submit any questions of substantial similarity for works that do not meet this threshold to the jury at trial. To make this assessment, their expert Dr. Goldstein removed from each Asserted Work any material in quotation marks and any extraneous non-letter characters. SF1177. Dr. Goldstein then calculated the percentage of identical 3-gram (three-word sequences) in the Asserted Work contained in the training data record. SF1175-1176. The texts were then reviewed to determine whether they shared a 16-gram (a 16-word sequence). SF1178-1181.

Below are excerpts of comparisons of two Asserted Works with the training data copies, where red text is used to show spans of at least 4 consecutive words in common:



SF1182.



SF1183.

The Times and DNP provided a hard drive to this Court containing comparisons of the Asserted Works and the closest training data copies that meet the foregoing conditions. *See* Maisel Declaration, Exhibits A and B (Rule 1006 Summaries that link to said comparisons); SF1175-1181. The Times and DNP seek summary judgment for acquisition, training, and distribution-based infringement as follows:

CIR seeks summary judgment based on training for [REDACTED] of its registered Asserted Works meeting these conditions. SF1208. OpenAI also copied [REDACTED] of unregistered CIR works, and [REDACTED] of The Intercept's unregistered works, which is relevant to their DMCA claims. SF1207, 1209-1210, 1212.

For summary judgment based on acquisition, CIR asserts all [REDACTED] registered Asserted Works in a Common Crawl-derived training set whose URL matched that of the work. SF1190, 1202-1203. Presence in a Common Crawl training set means that OpenAI made an initial copy from Common Crawl. SF813, 823-24, 1190. Those initial copies are identical to the original. SF818-819, 1274-1275.

iii) ZD

In this motion, ZD seeks summary judgment with respect to copies used for training and acquisition only as to its Asserted Works that are either an Exact Match or Revised Version copy, defined as follows: (i) the training data record contains the identical text of the Asserted Work (excluding the first and last 50 characters and excluding any CMI removed) or directional word containment is equal to or greater than 95% ("Exact Match"); or (ii) directional word containment is equal to or greater than 80% and 1-gram Jaccard similarity greater than 50% ("Revised Version"). SF1213-17.

Below are two excerpts of comparisons of two ZD Asserted Works with samples from the training data copies, where red text is used to show matching text:





SF1218–1220.

ZD has provided a hard drive containing matches of the Asserted Works with the closest training data copies that meet the foregoing conditions. *See* Koonce Declaration, Exhibit A;

SF1213-1228. ZD seeks summary judgment for acquisition, training, and distribution-based infringement as follows:

Infringing Act	ZD Copies and Asserted Works
Training	total number of copies unique Asserted Works
OpenAI's scraping of webpages for WebText	total number of copies unique Asserted Works
Defendants' scraping of webpages for Project	copies unique Asserted Works
OpenAI's downloading of Common Crawl datasets	copies unique Asserted Works
OpenAI's distribution of	copies unique Asserted Works

SF1223.

2. Copilot's RAG Pipeline and Outputs

As explained in Section II.B.3 above, Copilot copies Plaintiffs' content in real time during the RAG process—first when it retrieves content to use as a RAG source, and again when RAG outputs are substantially similar to the source retrieved. Each retrieval and output independently infringed Plaintiffs' exclusive right of reproduction by creating a separate reproduction, distinct from the reproductions Defendants made in acquiring Plaintiffs' works and training their models. 17 U.S.C. § 106(1).

Cases in the Second Circuit suggest a modest bar for the degree of “substantial similarity” required for a second work to be deemed an infringing copy or derivative. In *AP v. Meltwater*, the

court found that excerpts from news articles infringed, including one excerpt “up to 300 characters (including white [*i.e.*, blank] space) from the opening text of the article or lede,” and the second a “Hit Sentence,” which is “approximately 140 characters (not including white spaces) ‘surrounding a single algorithmically chosen appearance of one of the customer’s matched searched keywords.’” 931 F. Supp. 2d 537, 545 (S.D.N.Y. 2013). In *Keizai Shimbun, Inc. v. Comline Bus. Data, Inc.*, 166 F.3d 65, 70 (2d Cir. 1999), abstracts of the copyrighted news articles were substantially similar to the underlying works because the defendant used “the same structure and organization, follow[ed] the same chronological and substantive grouping of facts, and result[ed] in the same conclusions or resolutions, and in addition often employ[ed] identical phraseology and word choice.” *Id.* at 70-72.4.

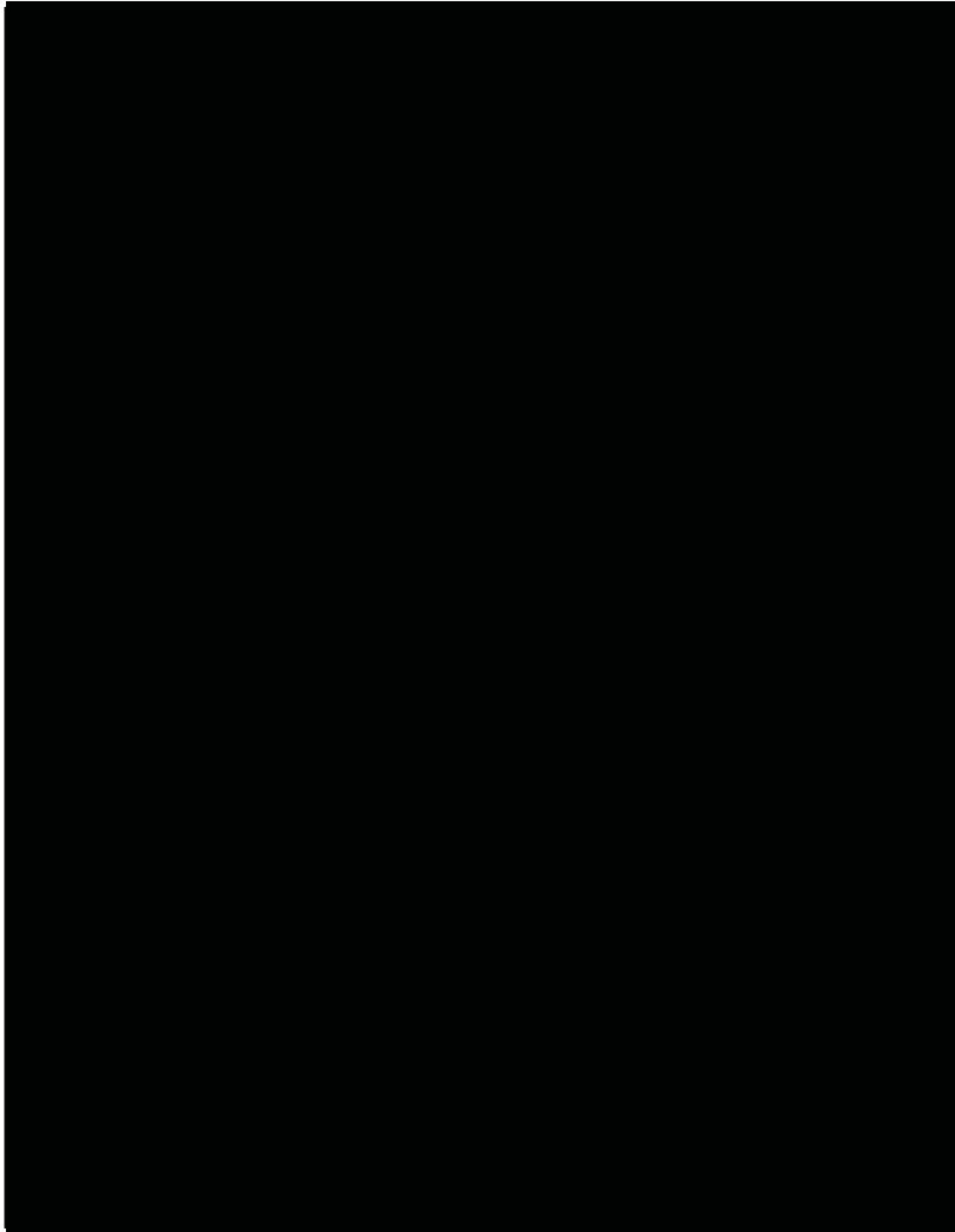
Here, Microsoft’s consumer Copilot conversation data, notwithstanding redactions and truncations, shows that Microsoft’s RAG process infringed Plaintiffs’ Asserted Works by copying them to ground Copilot’s responses, and then output verbatim portions or derivatives of the Asserted Works that closely follow the original.

Plaintiffs seek summary judgment with respect to an exemplary set of Asserted Works where there is no factual dispute that Microsoft made infringing copies in the operation of Copilot.

i) The Times and DNP

The Times and DNP move for summary judgment as to a sample of [REDACTED] and [REDACTED] Asserted Works, respectively, that Microsoft infringed in two distinct ways: (1) by copying substantial, verbatim portions of the works as part of the RAG process, and (2) by reproducing substantial verbatim portions or derivatives of the works to user. Plaintiffs reserve the remaining infringing Copilot copies for trial.

For illustration, in the one conversation Copilot retrieved more than [REDACTED] words from The Times's article "Trump and Allies Forge Plans to Increase Presidential Power in 2025." SF1239.



In the same conversation, Copilot output around [REDACTED] from that article to a user. SF1239.



In another conversation, Copilot retrieved around [REDACTED] words from the Denver Post article “Remains of 230 Native Americans still held in Colorado museums despite decades-old law ordering their return.” SF1240.



In this same conversation, Copilot output nearly all of the retrieved text to the user. SF1240.



Comparisons of The Times's and DNP's Asserted Works to the Copilot conversations are provided in Exhibits O and P, respectively, to the concurrently filed declaration of Plaintiffs' expert Dr. Tom Goldstein.

ii) CIR

CIR seeks summary judgment as to [REDACTED] registered Asserted Works that Microsoft copied either in its RAG data or in its output to users. SF1253. CIR's expert extrapolated these copies beyond Microsoft's sample, but because that extrapolation is disputed, CIR reserves the issue for trial. Two examples:

First, Copilot copied [REDACTED] consecutive verbatim words from a Mother Jones article, highlighted in yellow, as part of its RAG process. SF1255.



In another conversation, the user asks Copilot to [REDACTED] from a Mother Jones article. In response, Copilot outputted [REDACTED] verbatim words to a user. SF1256.



3. Defendants' Volitional Conduct

Defendants are responsible for their infringement. In this respect, the relevant question is **who** made the copy. Defendants did not merely “own[] a machine used by others to make illegal copies,” *Cartoon Network LP, LLLP v. CSC Holdings, Inc.*, 536 F.3d 121, 130 (2d Cir. 2008). As explained in Section II.B, *supra*, Defendants selected and copied the Asserted Works at each stage, bypassing Plaintiffs’ paywalls and other protective systems to select specific works for their datasets, models, and RAG indexes. *See Fox News Network, LLC v. Tveyes, Inc.*, 883 F.3d 169, 181 (2d Cir. 2018) (service that decided what content to make available to users engaged in volitional conduct).

IV. Defendants’ fair use defenses fail as a matter of law.

There is no precedent for Defendants’ claim that their unauthorized copying of millions of news articles that they could have licensed to train and operate their commercial AI products is “fair use.” Fair use is an affirmative defense that “allows the public to draw upon copyrighted materials without the permission of the copyright holder in certain circumstances” when “**necessary** to fulfill copyright’s very purpose, to promote the Progress of Science and useful Arts.” *HathiTrust*, 755 F.3d at 95 (cleaned up) (emphasis added). The four statutory factors are:

- (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 fair use provision . . . requires an analysis of the specific ‘use’ of a copyrighted work that is alleged to be an ‘infringement’” as “[t]he same copying may be fair when used for one purpose but not another.” *Warhol*, 598 U.S. at 533 (cleaned up); *see also Google Books*, 804 F.3d at 214-230 (2d Cir. 2015) (separately analyzing digitization of copyrighted works, creation of search functionality, display of snippets, and distribution of digital copies). Defendants have made unauthorized copies of Plaintiffs’ works for each of the distinct uses of (1) scraping (acquisition), (2) training LLMs that can generate news-like content, (3) RAG, (4) in outputs, and (5) by [REDACTED] [REDACTED]. Undisputed facts weigh against fair use on each factor and for each use.

First, the purpose and character of each use is to exploit the works for their intrinsic value and offer commercial products that supplant those works. Second, the nature of the works is expressive as a matter of law. Third, taking [REDACTED] of works in their entirety is not justified because Defendants themselves have taken the position that it was unnecessary to copy those works at all to achieve any ends they (incorrectly) argue are transformative. SF863-67. Fourth, Defendants’ products not only substitute for and harm the value of the works in their primary markets; they also usurp the market for AI licensing—a market that Plaintiffs would not just be likely to develop, but that many **have** developed and **are developing** in partnership with responsible AI vendors.

A. Factor 1: The purposes and character of the uses at issue are substitutive and commercial.

Under the first factor “courts consider two sub-factors: (i) the extent to which the secondary use is transformative, and (ii) whether the secondary use is commercial in nature.” *Hachette Book Group v. Internet Archive*, 115 F.4th 163, 179 (2d Cir. 2024). “[A] use does not become transformative by making an invaluable contribution to the progress of science and cultivation of the arts.” *HathiTrust*, 755 F.3d at 96 (quotation omitted). Rather, “a transformative work is one

that serves a new and different function from the original work and is not a substitute for it.” *Id.* “If an original work and a secondary use share the same or highly similar purposes, and the secondary use is of a commercial nature, the first factor is likely to weigh against fair use, absent some other justification for copying.” *Warhol*, 598 U.S. at 532–33. “Whether the purpose and character of a use weighs in favor of fair use is ... an objective inquiry into what use was made, *i.e.*, what the user does with the original work” compared with the purposes for which the original work was created. *Id.* at 545.

The first fair use factor further “weighs in favor of plaintiffs” where a defendant “knew that ... access to the [work] was unauthorized or was derived from a violation of law or breach of duty” and “a defendant could have acquired the copyrighted [work] legitimately.” *NXIVM Corp. v. Ross Inst.*, 364 F.3d 471, 477 (2d Cir. 2004). Thus an “evasive motive” weighs against a finding of fair use, *MCA, Inc. v. Wilson*, 677 F.2d 180, 183 (2d Cir. 1981), as does obtaining access to a work through “violat[ion]” of “industry standards of ethical behavior” or “misrepresentation,” *Roy Exp. Co. Establishment of Vaduz v. CBS*, 503 F. Supp. 1137, 1146 (S.D.N.Y. 1980). *See Harper & Row Publishers, Inc. v. Nation Enters.*, 471 U.S. 539, 563 (“knowingly exploit[ing] a purloined manuscript” not fair use).⁶

1. The substitutive purposes of defendants’ copying weigh against fair use.

Defendants’ top executives admitted that the generative AI (“GenAI”) products they build and operate by copying news [REDACTED]. Each of the uses that Defendants have made of Plaintiffs’ works is implicated in this overall purpose. As noted, Microsoft’s CEO testified under

⁶ Subsequent cases questioning the relevance to the first fair use factor of an infringer’s state of mind in, for example, failing to first seek permission, *see, e.g., Cambell*, 510 U.S. at 586 n.18, have neither addressed deceptive or unethical conduct in obtaining access to works nor overruled *Harper & Row* or the Second Circuit’s holding in *NXIVM* that failing to consider such evidence is error. *See* 364 F.3d at 478.

oath that conversations with chatbots like Copilot [REDACTED] SF1432, 1510, and OpenAI senior executive Nick Turley wrote that [REDACTED] SF1473. Microsoft executive Dr. Brent Hecht, in a document he wrote soon after this lawsuit was filed, explained that Defendants' AI systems [REDACTED] [REDACTED] [REDACTED] SF1672.

OpenAI Policy Director Jack Clark similarly wrote: [REDACTED]

[REDACTED] SF1677. OpenAI internal documents characterize ChatGPT as [REDACTED] SF1500, and brag that ChatGPT provides [REDACTED] [REDACTED] SF1526.

Speaking about the capabilities of models trained on [REDACTED]

[REDACTED], OpenAI executive Greg Brockman told his colleagues: [REDACTED]

[REDACTED] SF1425. Mr. Brockman also said that OpenAI's model [REDACTED]

[REDACTED] SF1427. OpenAI further acknowledges that its ChatGPT models [REDACTED] SF872, 1425.

In a presentation, Dario Amodei, the current CEO of Anthropic and then a top researcher at

OpenAI, listed [REDACTED] as one of GPT-3's [REDACTED] and he identified a [REDACTED]
[REDACTED] SF1429-30.

With respect to RAG, Dr. Hecht agreed that: [REDACTED]
[REDACTED] SF1436. Microsoft's Copilot home page says that "instead of clicking through links [e.g., to Plaintiffs' websites] we [Copilot] can talk about whatever you're curious about." SF1524. OpenAI's Nick Turley [REDACTED]
[REDACTED]

[REDACTED] SF1525. *See also* SF1539 (third party study reporting that 87.78% of ChatGPT users visit no external websites during their search, compared to only 26.91% of Google users).

With respect to outputs that are substantially similar to training or grounding sources, courts have rejected claims that copying news articles to provide a product that substitutes for demand for news is fair use. *Associated Press v. Meltwater U.S. Holdings, Inc.*, for example, found that factor one weighed against a finding of fair use for a defendant that "use[d] its computer programs to automatically capture and republish designated segments of the text from news articles, without adding any commentary or insight" or "improv[ing] access to the complete, linked news story[.]" 931 F.Supp.2d 537, 552–54 (S.D.N.Y. 2013).

Thomson Reuters Enterprise Centre GmbH v. Ross Intelligence Inc., reached the same result even in the absence of similar outputs. 765 F. Supp. 3d 382 (D. Del. 2025). In copying Westlaw headnotes to train an AI legal-research tool that returned relevant judicial opinions in response to a user's legal query, the court found that the defendant, Ross, "was using Thomson Reuters's headnotes as AI data to create a legal research tool to compete with Westlaw." *Id.* at 398. This use was not transformative because it had no "further purpose or different character"

from Westlaw’s own headnotes and Key Number System, which serve the identical function of helping researchers find relevant case law. *Id.* at 397–99 (quoting *Warhol*, 598 U.S. at 529).

2. Defendants’ commercial purposes weigh against fair use.

“[A] commercial motivation on the part of the secondary user will weigh against” a finding of fair use under the first factor, “especially, as the Supreme Court suggested [in *Campbell*], when a persuasive transformative purpose is lacking.” *Google Books*, 804 F.3d at 219 (citing *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 579 (1994)). The purposes for which Defendants copied Plaintiffs’ works do not compare to any that controlling authorities have found to overcome this presumption under factor 1.

Scraping, training on, grounding on, outputting, and [REDACTED] copies or derivatives of copyrighted news articles do not “provide otherwise unavailable information about the originals,” as in *Google Books*, 804 F.3d at 215, much less “reimplement an interface [that] can further the development of computer programs,” as in *Oracle*, 593 U.S. at 31, or “use of some elements of a prior author’s composition to create a new one that, at least in part, comments on that author’s works,” as in *Campbell*, 510 U.S. at 580. Defendants’ uses simply consume the works to profit, respectively, from accumulating (scraping), approximating (training), emulating (grounding), reproducing (outputting), and selling (dealing in) their expression. A “consumptive use” of a work “for a purpose that exploits its intrinsic expressive value [] fails the test” of transformation. 4 Nimmer on Copyright § 13.05F; see *Twin Peaks Prods., Inc. v. Publications Int’l, Ltd.*, 996 F.2d 1366, 1375 (2d Cir. 1993) (“[W]e have [] rejected a fair use defense for works that could be characterized as ‘commercial exploitation.’”).

The direct commercial purposes of Defendants’ uses contrast with *Google Books*. In the product at issue there,

[n]o advertising is displayed to a user of the search function. Nor does Google

receive payment by reason of the searcher's use of Google's link to purchase the book....Google has no revenues flowing directly from its operation of the Google Books functions[.]

Google Books, 804 F.3d at 209, 218. OpenAI and Microsoft both charge for subscriptions to use, and for advertising in connection with, their products that exploit Plaintiffs' works. SF576-81, 607-11, 649, 652, 660. The fair use finding in *Google Books* "test[ed] the boundaries of fair use" by commercial actors. 804 F.3d at 206. The uses here vastly overstep them.

3. The character of Defendants' conduct weighs further against fair use.

The transgressive and deceptive nature of Defendants' efforts to acquire Plaintiffs' content confirms that their purpose in doing so was fundamentally to avoid the cost of acquiring lawful copies. OpenAI employees devised and implemented strategies to [REDACTED]. For example, when Nick Ryder informed Mr. Brockman about [REDACTED] to help with Mr. Brockman's [REDACTED] Mr. Brockman responded [REDACTED] SF791. OpenAI's corporate representative on the subject was unaware of [REDACTED]. SF463. OpenAI also hosts CustomGPTs designed to enable users of its products to access paywalled content without permission. SF588-596, 1015, 1468-70.

OpenAI's disregard of paywalls, terms of service, and other restrictions violated industry norms. Mr. Nadella testified that [REDACTED] and made clear that, if he [REDACTED] [REDACTED] SF794, 796. Mr. Nadella likewise agreed [REDACTED] SF795. But OpenAI's general practice when directly crawling and scraping data [REDACTED] SF521.

OpenAI acted improperly in other ways. [REDACTED]

[REDACTED]. SF1093-1154.

Microsoft violated industry norms by selling the Bing Index to OpenAI to use as training data knowing that Plaintiffs and others had only allowed Microsoft to crawl that data for purposes of search. SF689-711. And OpenAI will offer no evidence that it believed at the time that any of this conduct was justified by fair use. Dkt. 1276 (February 6, 2026, Order in Class case) at 7-8; SF1821 (counsel for OpenAI agreeing by email that the same applies to the News cases).

B. Factor 2: The expressive nature of the asserted works disfavors fair use.

The second fair use factor is “the nature of the copyrighted work.” 17 U.S.C. § 107(2). This 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. Factor two turns on whether the works in question—whether factual or fictional—“reflect ‘the author’s individualized expression.’” *Hachette*, 115 F.4th at 187 (quoting *Harper & Row*, 471 U.S. at 563) (holding that although the Copyright Act “does not protect facts or ideas set forth in a work... it **does** protect that author’s manner of expressing those facts and ideas”) (cleaned up) (original emphasis). Because the Asserted Works “contain original expression close to the core of intended copyright protection,” such as “the author’s compilation of those facts” into an expressive form, they are “‘of the type that the copyright laws value and seek to protect,’ [and] the second fair use factor favors.” *Id.* (quoting *HathiTrust*, 755 F.3d at 98).

C. Factor 3: The amount and substantiality of the portions used are unreasonable in relation to the purpose of the copying.

The third fair use factor focuses on “the amount and substantiality of the portion used in relation to the copyrighted work as a whole.” 17 U.S.C. § 107(3). Copying large volumes of

copyrighted works in their entirety can only be justified as fair use when and to the extent “doing so [is] necessary to achieve a transformative, secondary purpose.” *Hachette*, 115 F.4th at 188. Defendants indisputably copied all the asserted works substantially in their entirety during scraping, training, and [REDACTED] with each other. Section II.B, *supra*. Yet Defendants’ experts affirmatively assert that Defendants [REDACTED]

[REDACTED].

OpenAI’s technical expert Dr. Chris Callison-Burch states that [REDACTED]

[REDACTED]

[REDACTED] SF863. OpenAI expert Dr. Avi Goldfarb similarly opined that

[REDACTED] SF864. Microsoft’s technical expert Dr.

John Lafferty made the same assertion: [REDACTED]

[REDACTED]

[REDACTED]

[REDACTED] SF866. Dr. Lafferty elaborated in his deposition that news content [REDACTED]

[REDACTED]

[REDACTED]

[REDACTED] SF867. The admitted lack of connection between the [REDACTED]

of Plaintiffs’ copied works and Defendants’ purported transformative purpose⁷ of creating LLMs that can generalize makes this factor weigh strongly in Plaintiffs’ favors.

⁷ There is, however, also no dispute that training and grounding on at least some Plaintiffs’ works, like, for example, the New York Times’s works that OpenAI [REDACTED] are highly valuable for the non-transformative commercial purpose of being able to respond to user queries seeking news. SF869-873.

D. Factor 4: Defendants cannot show lack of harm to the market for and value of Plaintiffs’ works.

The fourth factor requires considering “the effect of the use upon the potential market for or value of the copyrighted work.” 17 U.S.C. § 107(4). “Since fair use is an affirmative defense, its proponent ... carr[ies] the burden of demonstrating fair use with[] favorable evidence about relevant markets.” *Campbell*, 510 U.S. at 590. Such evidence must prove not only the lack of any “likely effect on the market for the original,” but also “for potential derivative uses ... that creators of original works would in general develop or license others to develop.” *Id.* at 590, 592; *Hachette*, 115 F.4th at 194 (“[T]he ultimate burden of proving that the secondary use does not compete in the relevant market is appropriately borne by the party asserting the defense: the secondary user.”).

“[G]iven the fourth factor’s importance, it’s easy to imagine a situation in which a secondary use is highly transformative but the secondary user nonetheless loses on fair use because allowing people to engage in that kind of use would have too great an effect on the market for the original work.” *Kadrey*, 788 F. Supp. 3d at 1051. Indeed, in *Campbell*, the Supreme Court remanded for further consideration of the fourth factor despite finding that the purpose of engaging in parody was highly transformative under the first factor. 510 U.S. at 593–94.

Factor four accounts for “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.” *Hachette*, 115 F.4th 189 (internal quotation marks omitted). It is “concerned with the **category** of a defendant’s conduct, not merely the specific instances of copying.” *Texaco Inc.*, 60 F.3d 913 at 927 n.12 (emphasis added); *see also* 4 Melville B. Nimmer & David Nimmer, NIMMER ON COPYRIGHT § 13F.08 (2023) (“[F]actor four is expansive enough to weigh the impact of hypothetical uses beyond defendant’s own.”); *Nicklen v. Sinclair Broad. Grp., Inc.*, 551 F. Supp. 3d 188, 198 (S.D.N.Y. 2021) (concluding that factor four weighed against

fair use because “widespread adoption” of defendants’ misconduct by parties other than the defendants could “overtake” the licensing market for plaintiff’s work).

To counter a defendant’s evidence purporting to show lack of harm, a rightsholder need not offer “empirical data” showing actual effects. *Hachette*, 115 F.4th at 193. Rather, courts “routinely rely on ... logical inferences where appropriate in assessing the fourth fair use factor.” *Id.* (collecting cases). In *Warhol*, for example, the Second Circuit held that factor four favored the photographer because of the harm that would occur if “the sort of copying in which Warhol engaged were to become a widespread practice.” *Andy Warhol Found. for Visual Arts, Inc. v. Goldsmith (“Warhol I”)*, 11 F.4th 26, 50 (2d Cir. 2021).⁸ It found that excusing unauthorized copying like Warhol’s “would effectively destroy [a] broader market” for licensing, “as, if artists could use such images for free, there would be little or no reason to pay for them.” *Id.* (internal quotations and alterations omitted). No empirical evidence was necessary for the Second Circuit to reach that commonsense conclusion. Because the market harm was “self-evident,” the fourth factor favored Goldsmith. *Id.* at 50-51.

Here, at least three types of market harm cut against fair use.

First, is the substitutive harm to the value of the Asserted Works themselves from each of the infringing uses. Where the defendant “brings 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 rather than the original,” factor four disfavors fair use. *Hachette*, 115 F.4th at 189.

⁸ The Supreme Court’s subsequent affirmance in that case related only to factor one and did not disturb the Second Circuit’s treatment of factor four. *Warhol*, 598 U.S. at 525 (2023) (“Although the Court of Appeals analyzed each fair use factor, the only question before this Court is whether the court below correctly held that the first factor...weighs in Goldsmith’s favor.”)

Hachette provides the classic example of this type of harm. There, the defendant offered “effectively the same product as” rightsholders—books—and made them available, in full, at no cost to consumers or libraries. *Id.* at 190. As the court put it, it is “difficult to compete with free.” *Id.* But complete, identical, copies are not required. *Google Books*, for example, found that copying to enable search for book contents was fair use only because the digital library’s “snippet view” produced “discontinuous, tiny fragments” of books that did not give searchers access to “effectively competing substitutes” for the books that Google had copied. 804 F.3d at 224. As the Second Circuit later noted, “the judgment [in *Google Books*] would likely have come out the other way if the copying had enabled the public to read substantial parts of the copied books.” *Romanova*, 138 F.4th at 117 n.9. Thus, courts considering whether GenAI models trained on copyrighted text qualify for fair use recognize that where the end product “could be used to generate significant portions” of the copyrighted works, factor four would cut in favor of the rightsholders. *Kadrey*, 788 F. Supp. 3d at 1051 (citing *Google Books*, 804 F.3d at 224).

Second is the harm to licensing markets. “It is indisputable that, as a general matter, a copyright holder is entitled to demand a royalty for licensing others to use its copyrighted work ... and that the impact on potential licensing revenues is a proper subject for consideration in assessing the fourth factor.” *Texaco*, 60 F.3d at 929. And where the copyright holder can exploit the copyrighted work in multiple licensing markets, the court should consider the effects in all those markets. *See, e.g., Warhol I*, 11 F.4th at 49–50 (separately analyzing harm to the market for magazine licensing of photographs, and harm to the market for licensing of “stylized derivatives” of photographs). The cognizable licensing markets include not just the markets in which the rightsholder currently participates, but also those that it might yet enter. *See, e.g., Castle Rock Ent., Inc. v. Carol Pub. Grp. Inc.*, 150 F.3d 132, 145–46 (2d Cir. 1998); *see also Thomson Reuters*

Enter. Ctr. GMBH v. Ross Intel. Inc., 765 F. Supp. 3d 382, 400 (D. Del. 2025) (“I must consider not only current markets but also potential derivative ones ‘that creators of original works would in general develop or license others to develop.’” (quoting *Campbell*, 510 U.S. at 592)).

Markets that may be considered under factor four must be “traditional, reasonable, or likely to be developed” so as to guard against the “vice of circular reasoning” that would result if courts could find harm simply because a copyright holder was not paid for a use that would not typically require a licensing fee. *Texaco*, 60 F.3d at 930–31. Thus, factor four disfavors fair use where there is a “workable” and “viable” market to obtain a license. *Texaco*, 60 F.3d at 930; *Tveyes*, 883 F.3d at 180 (factor four disfavored fair use because defendant usurped a “plausibly exploitable market” for access to copyrighted television content).

Third is the harm to the market for the same type of content as the Asserted Works—both the market for the publisher’s own body of works and the market for the broader category (here, news articles)—from unrestricted and widespread adoption of Defendants’ conduct. *See Hachette*, 115 F.4th at 193. The statute confirms that factor four reaches beyond harm from unauthorized copies. Not only does it direct courts to consider the use’s effect “upon the potential *market for*...the copyrighted work,” but it also directs courts to consider “the effect of the use upon the...*value of* the copyrighted work.” 17 U.S.C. § 107(4) (emphases added). The phrase “value of” has a broader meaning than “market for”—indeed, reading the two synonymously would violate the principle that words in a statute are to be given independent meaning, particularly when separated by the word “or.” *See United States v. Woods*, 571 U.S. 31, 45 (2013); Jane C. Ginsburg, *Fair Use Fair Factor Four Revisited: Valuing the “Value of the Copyrighted Work,”* 67 J. COPYRIGHT SOC’Y U.S.A. 19 (2020) (“[T]he statute’s designation of ‘the value of the copyrighted work’ identifies independent kinds of harm and entails considerations distinct from market substitution.”). A use

that devalues an entire genre of works necessarily affects the “value of the copyrighted work” within that genre as well.

The Second Circuit’s analysis in *Warhol I* reflects the breadth of harms captured by factor four. There, the Court did not just consider the harm to the market for Goldsmith’s photographs of Prince. Instead, it examined the harm to the “broader market”—the “market to license photographs of musicians, *such as* the Goldsmith Photograph, to serve as the basis of a stylized derivative image[.]” 11 F.4th at 50. By accounting for the harm in this market, the court guarded against the “risk[.]” of “disincentivizing artists from producing new work by decreasing its value—the precise evil against which copyright law is designed to guard.” *Id.*

Courts considering fair use in the AI training context have held that this type of harm plays a critical role in the factor four analysis. As *Kadrey* put it:

This case, unlike any of those cases [such as *Google Books* and *Perfect 10, Inc. v. Amazon.com, Inc.*], involves a technology that can generate literally millions of secondary works, with a miniscule fraction of the time and creativity used to create the original works it was trained on. No other use—whether it’s the creation of a single secondary work or the creation of other digital tools—has anything near the potential to flood the market with competing works the way that LLM training does....

Courts 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. Indeed, it seems likely that market dilution will often cause plaintiffs to decisively win the fourth factor—and thus win the fair use question overall—in cases like this.

Kadrey, 788 F. Supp. 3d at 1055; *see also* United States Copyright Office, COPYRIGHT AND ARTIFICIAL INTELLIGENCE, PART 3: GENERATIVE AI TRAINING (May 2025) at 65–66 (“the fourth factor should not be read so narrowly,” and should account for harm that “stem[s] from AI models’ generation of material stylistically similar to works in their training data...[e]ven when the output is not substantially similar to a specific underlying work[.]”).

Indeed, *Kadrey* highlighted how potent this argument may be in the context of the news market. As Judge Chhabria explained, “markets for certain types of works (*like news articles*) might be even more vulnerable to indirect competition from AI outputs” than others, “present[ing] even stronger arguments against fair use.” *Kadrey*, 788 F. Supp. 3d at 1059 (emphasis added). Or, as Judge Chhabria explained elsewhere, “[a]n LLM that could generate accurate information about current events might be expected to greatly harm the print news market.” *Id.* at 1053. Beyond this, LLMs may also generate content that resembles news content, but in fact contains *inaccurate* information about current events—a phenomenon potentially even more harmful to the market for news. *See infra* Section IV.D.1.iv. They may even falsely attribute such content to news publishers. However considered, Judge Chhabria’s recognition that harm to the market for news may cause substantial, cognizable harm under factor four is correct and borne out by the evidence in this case.

Courts also sometimes consider, as part of the factor-four analysis, the “public benefits the copying will likely produce.” *Google*, 593 U.S. at 35. Importantly, however, the “public benefit” analysis is not a free-floating assessment of all theoretical benefits that might be derived from the copying. Instead, the public benefits that are cognizable are those that are related to the goals of copyright—specifically, the “creative production of new expression.” *Id.* at 35–36. And those benefits must be balanced against copyright-related losses, both to the rightsholders and to society more generally.

In *Hachette*, for example, the Second Circuit rejected the defendant’s invocation of “public benefits” from providing a “free digital library” of physical books because those benefits were “outweighed not only by harm to Publishers and authors but also by the long-term detriments society may suffer if [defendant’s] infringing use were allowed to continue.” 115 F.4th at 196. It found that the “long-term consequences” following from “unrestricted and widespread” use of

authors’ content would leave “little motivation to produce new works,” resulting in a “dearth of creative activity [that] would undoubtedly negatively impact the public,” a “reality that the Copyright Act seeks to avoid.” *Id.* at 195; *see also Associated Press v. Meltwater U.S. Holdings, Inc.*, 931 F. Supp. 2d 537, 553 (S.D.N.Y. 2013) (“Permitting Meltwater to take the fruit of AP’s labor for its own profit, without compensating AP, injures AP’s ability to perform this essential function of democracy”). Such an outcome would not only “decimate Publishers’ markets for the Works in Suit across formats,” *Hachette*, 115 F.4th at 195, but set off the “doom loop” that a Microsoft executive has said would imperil the future of AI itself.

1. Defendants’ generative AI models and products substitute for news.

Defendants’ products generate verbatim copies, summaries, and mimicry of Plaintiffs’ works because Defendants trained and grounded on those works. Users submit [REDACTED] of queries to Defendants’ products seeking news, and unlike traditional search engines, the AI-generated outputs provide substitutive answers that reduce the need for users to visit Plaintiffs’ websites and flood the market with low-quality material that undermines the business model of news publishers.

i) Defendants’ products output news content.

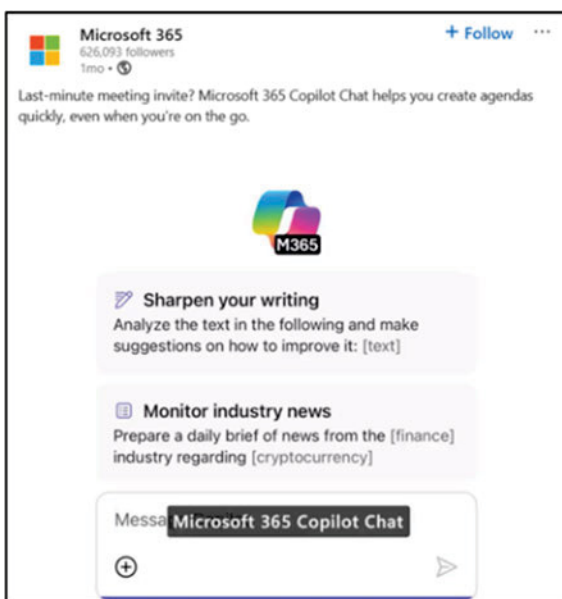
This case involves exactly the situation hypothesized in *Romanova* where “the copying ... enable[s] the public to read substantial parts of the copied [works].” 138 F.4th at 117 n.9. As demonstrated above in Section II.B, *supra*, Defendants’ models and products can and do produce outputs that contain verbatim copies of Plaintiffs’ asserted works as well as paraphrased copies and detailed summaries that are substantially similar to the asserted works themselves. SF942-970, 1053-1054, 1057-1074, 1078-1092, 1243-1256, 1422-1436.

Even where ChatGPT and Copilot outputs are not verbatim copies or summaries of Plaintiffs’ asserted works, training and grounding on those works is what enables models to substitute for news by reproducing the selection, coordination, and arrangement of the information

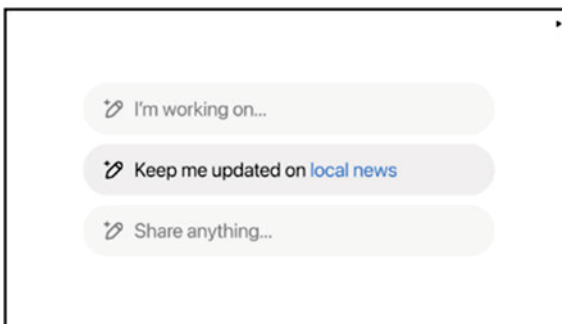
those works contain while emulating their expressive form and style. SF1425-36. This exploitation of copyrightable elements of Plaintiffs’ articles even without obvious regurgitation is why OpenAI’s Head of ChatGPT, concluded that [REDACTED]

[REDACTED] SF1473.

Users frequently seek news from Defendants’ products. SF1479-91. Their marketing confirms that this is a key use for ChatGPT and Copilot. Microsoft 365 Copilot Chat promoted its ability to “Monitor industry news” and “prepare a daily brief on news...”:



SF1492. Similarly, OpenAI’s ChatGPT Pulse offers to keep users “updated on local news”:



SF1496.

Defendants’ internal emails and documents also tout these capabilities. SF1423-430. For example, OpenAI noted that its models were [REDACTED] SF1426-27. As early as 2020, OpenAI listed [REDACTED] SF1429. OpenAI also identified [REDACTED] SF1430. A study by multiple OpenAI authors confirms that [REDACTED] accounts for [REDACTED] of ChatGPT’s usage in 2025. SF1522. Local news, an important aspect of the DNP’s business, has been described by OpenAI as [REDACTED] SF1640. OpenAI’s expert admits that their products are [REDACTED] SF1487. Custom GPTs like “News Summarizer” and “Remove Paywall” confirm that this is a nontrivial use case. SF1470. Defendants do not dispute that the reason ChatGPT and Copilot are [REDACTED] SF1428.

ii) Defendants’ products do not drive traffic to sources.

The products are substitutive because they [REDACTED] SF1462. As Microsoft acknowledged: [REDACTED] [REDACTED] [REDACTED] SF1476. That’s why OpenAI understood that its systems posed [REDACTED] SF1466. As an OpenAI software engineer stated, [REDACTED] SF1475.

⁹“Search engine response page”—*i.e.*, a page of links to search results.

Defendants also acknowledge that ChatGPT, Copilot, and GenAI models aim to supplant traditional search. Sam Altman has called ChatGPT “a Google replacement.” SF1519. Microsoft’s survey expert reported that [REDACTED] of Copilot survey respondents would [REDACTED] [REDACTED] SF1515. The court in the government’s search monopoly case against Google rejected strong remedies because of emerging competition from GenAI. SF1518.

The replacement of search with GenAI harms publishers by producing **drastically lower** click-through rates (“CTRs”).¹⁰ SF1530-1537. As explained above, Plaintiffs are not required to prove harm empirically: common sense suffices. But Microsoft’s representative CTR data shows that the overall click-through rate reduction between Bing Chat and Bing Web Search was [REDACTED] [REDACTED] for The Times’s websites, [REDACTED] for DNP’s websites, and [REDACTED] for ZD websites. SF1536. Such reductions in CTR clearly show substitution. *See AP v. Meltwater*, 931 F.Supp.2d at 554-55 (“Instead of driving subscribers to third-party websites,” minimal CTR shows “Meltwater News acts as a substitute for news sites...”).

Defendants’ own experts acknowledge that [REDACTED] [REDACTED] As Microsoft’s economist expert Dr. Tucker explained:

[REDACTED]

¹⁰ “Click-through rate” refers to the percentage of clicks from users to a given URL relative to the total impressions, which are the appearances of that URL in a search results page or within a chatbot’s response. SF1530.

SF1523. Defendants aren't shy about this fact. They advertise "fast, timely answers" instead of going to a search engine, SF570, 1526, and "instead of clicking through links, we can talk about whatever you're curious about." SF1524.

Defendants acknowledge the predictable consequences of this design. Per Microsoft, the

[REDACTED]

[REDACTED] SF1798. They [REDACTED]

[REDACTED]

[REDACTED]. SF1452, 1677. There is a [REDACTED]

[REDACTED]

[REDACTED] SF1467. [REDACTED]

SF1461.

iii) Defendants' products generate pink slime.

Another way that Defendants' accused products substitute for news is by enabling the generation of high volumes of "pink slime," which is unreliable low-quality content that is often either plagiarized or remixed from legitimate news sources. SF1683-1684. Defendants' accused products have made pink slime news content farms cheaper, easier to make, and harder to detect. SF1686. Using OpenAI's published API pricing, it would cost roughly \$6,800 to generate one million news-style articles averaging 500 words each – and would require no human author, editor, or reporter to do so. SF1491.

The proliferation of GenAI pink slime news content is harmful to journalism because it undermines the business model of news publishers by diluting the market with a flood of low-quality inauthentic news material, publishing plagiarized content, and siphoning away readers and advertising money. SF1703, 1706-1709. For example, Prism News operated 200 AI-generated publications posing as local newsrooms and hobby sites, with only four employees. SF1710.

Rather than conduct any form of reporting, Prism used GenAI to summarize existing articles and repackaging them as “new” stories competing in the original articles’ markets. SF1711. Prism News is but one example of why the “concept of market dilution becomes highly relevant” to the consideration of fair use when GenAI tools are at issue. *Kadrey*, 788 F.Supp.3d at 1055.

Defendants [REDACTED]

[REDACTED] and that GenAI pink slime news articles were often indistinguishable to readers from original content. SF1683-1712.

iv) Defendants’ products diminish demand for Plaintiffs’ content.

Plaintiffs have already begun to experience the effects of Defendants’ products on the market for news. Surveys show that Plaintiffs’ subscribers are aware of and starting to use GenAI tools. SF1579-1595. Former DNP digital subscribers are [REDACTED] more likely and former print subscribers are [REDACTED] more likely to use GenAI tools than active digital and active print subscribers respectively. SF1594. Among Times subscribers who were surveyed who used or paid for ChatGPT for news, [REDACTED] responded that their use of ChatGPT “means I no longer need news from the New York Times at all.” SF1581.

DNP has experienced declining search referrals and revenue since ChatGPT launched in 2022. SF1609-1625.

ZD has experienced the same pattern. ZD invests heavily in search-engine optimization, advertising, and other efforts designed to increase monetizable traffic to its properties, but those monetization efforts have been undermined by declines in traffic and revenues attributable to Generative AI answer engines. SF1626. Cloudflare’s CEO cited data showing the ratio of webpages scraped to visitors sent: in 2015, Google scraped two pages for every visitor (2:1); by June 2025, that ratio was 18:1 for Google and 1,500:1 for OpenAI. SF1540.

CIR and The Intercept, nonprofit news organizations, likewise depend on search traffic to acquire donations and grants to support their mission. SF104-110. That traffic has shrunk because of GenAI products and is likely to shrink further. SF1627-1628.

Further, when Defendants' use captures that scarce audience and attention using Plaintiffs' own content, there is not only less demand for Plaintiffs' content in their own platforms, but also less demand within their paying licensees' distribution channels and platforms. SF1629-1634.

2. Defendants' uses supersede feasible licensing markets.

Defendants' uses of Plaintiffs' copyrighted works also harm the existing and likely market for licensing news content for use in LLM training and RAG. Defendants took Plaintiffs' works for free even though they pay license fees to other publishers for the same kind of content and for the same uses. SF147-206, 207-228. This is a well-recognized theory of harm. *Campbell*, 510 U.S. at 590, 592; *Hachette*, 115 F.4th at 194.

The market for licensing data for LLM training and RAG already exists.¹¹ SF135-141. In fact, Defendants helped build it. SF147-206, 207-228. OpenAI has entered into at least [REDACTED]

[REDACTED]. SF141, 147-151. Annual payments under the [REDACTED] vary widely, from [REDACTED]

[REDACTED] SF149. [REDACTED]

[REDACTED] SF150. Similarly, Microsoft has entered into at least [REDACTED]

¹¹ Even Defendants' experts acknowledge [REDACTED] which exceeds the "likely to be developed" standard in *Texaco*, 60 F.3d at 930. See SF136 (Microsoft's expert Dr. Tucker admitting that these [REDACTED]; SF138 (Microsoft's economic expert Dr. Rao using [REDACTED]

[REDACTED]

[REDACTED] SF207. [REDACTED]

[REDACTED] SF146.

Third parties are paying to license this type of content for use in GenAI tools. For example, Amazon has [REDACTED]

[REDACTED]

[REDACTED]

SF247-273. Additional GenAI providers, including Google, Meta, Perplexity, ProRata.ai, and Mistral AI, have entered agreements with news publishers. SF274. Companies such as TollBit, Cloudflare, Credent, ProRata, ScalePost, Human Native AI, Copyright Clearance Center, Dappier, Defined.ai, Goodrights, Valyu, Personal Digital Spaces, and Calliope Networks connect publishers with GenAI companies to monetize digital data. SF275-290. Yet rather than pay Plaintiffs royalties to access their copyrighted [REDACTED] SF1651, as they have repeatedly done for other publishers, SF147-206, 207-228, Defendants chose to [REDACTED] Plaintiffs' copyrighted content for free. SF1437.

3. Widespread and unrestricted adoption of Defendants' conduct would severely harm the market for news along with AI itself.

There is no doubt about "what would happen" if Defendants' conduct were to become widespread and unrestricted—because it already is happening. One effect of ChatGPT's release in November 2022 was to spark other companies like Google to fast-track their own GenAI products, including, most significantly, Google's AI Overviews. SF1779. "Today," Microsoft admits, [REDACTED]

[REDACTED]

[REDACTED]

[REDACTED] SF1778.

Defendants' own experts admit [REDACTED]

[REDACTED] SF1781-1788. OpenAI's economic expert Dr. Stamm opined that [REDACTED] SF1782.

Another OpenAI economic expert, Dr. Goldfarb, opined [REDACTED]

[REDACTED]

[REDACTED] SF1784, 1786. Dr. Goldfarb also opined [REDACTED]

[REDACTED]. SF1785. Dr. Sinnreich, OpenAI's media expert, opined based on a 2026 Reuters Institute analysis that [REDACTED]

[REDACTED]

[REDACTED]

[REDACTED] SF1787. Dr. Sinnreich also admitted that [REDACTED]

[REDACTED] SF1788.

The harm flowing from unrestricted and widespread adoption has already begun. OpenAI understood that its technology [REDACTED].

SF1466. Microsoft has observed that [REDACTED]

[REDACTED]

[REDACTED]

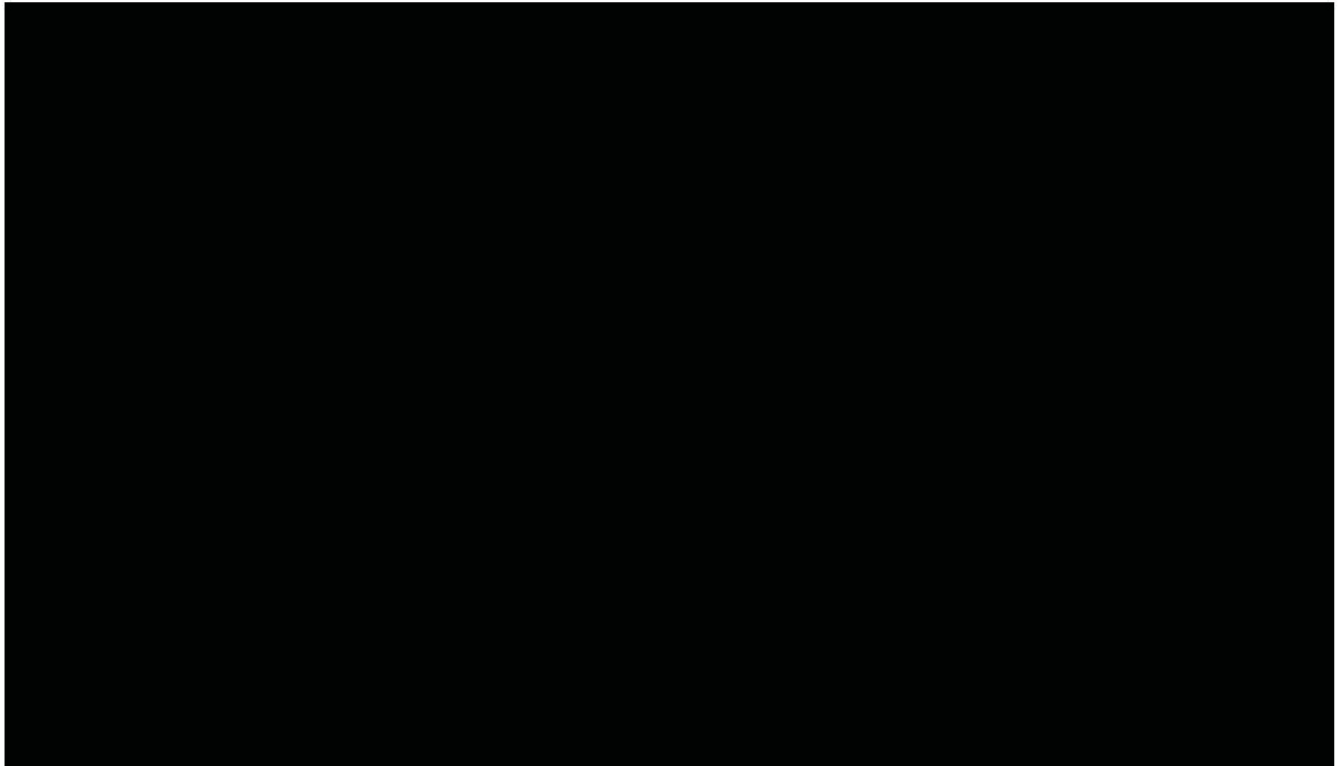
[REDACTED]

[REDACTED] SF1672.

Yet AI companies remain powerless to break out of this [REDACTED] because, while the industry as a whole would benefit if every company paid to sustain the continued production of the creative works their technology depends on, each individual company is better off taking

content for free while others pay. Finding that copying news for AI is not fair use would solve this prisoners' dilemma by putting all AI companies, OpenAI and Microsoft included, on an even footing. As Microsoft's Dr. Glen Weyl put it, compensating creators is [REDACTED]

[REDACTED] SF1657. An illustration from a Microsoft document that is worth a thousand words shows why:



SF1817.

E. None of the uses at issue in this case are fair.

To sum up, all statutory factors weigh against Defendants' affirmative defense of fair use for each copy that Defendants made of Plaintiffs' articles.

- **Acquisition:** Deceptively scraping expressive works, often from behind paywalls and in violation of terms of service and other restrictions or obtaining them from

unlicensed third-party scrapers to curate a general-purpose repository without paying the usual costs of acquisition has no fair use justification.

- **Training:** Copying Plaintiffs’ articles to train models was not fair use because doing so was admittedly unnecessary to realize any general capabilities but was done instead to serve commercially valuable news-seeking use cases that substitute for those articles.
- **Grounding:** Microsoft’s copying of Plaintiffs’ articles to respond to user queries self-evidently substitutes for the articles it copied.
- **Output:** Microsoft’s delivery of copies or recognizable derivatives of copyrighted works to the public in Copilot outputs that do not generate significant referral traffic lacks any fair use purpose.
- [REDACTED] Defendants’ copying of Plaintiffs’ content to [REDACTED] it with each other is just as obviously not fair use.

V. The Court should grant summary judgment to plaintiffs on the first three elements of their DMCA claims.

The DMCA was enacted “to strengthen copyright protection in the digital age,” “combat copyright piracy,” and provide “broad protections to copyright owners.” *N.Y. Times Co. v. Microsoft Corp.*, 777 F. Supp. 3d 283, 312 (S.D.N.Y. 2025); *Mango v. BuzzFeed, Inc.*, 970 F.3d 167, 172 n.2 (2d Cir. 2020). A violation of section 1202(b)(1) requires: (1) conveyance of copyright management information (“CMI”) in connection with a copy of a work, (2) removal or alteration of CMI without authority of the copyright owner or the law, (3) that the removal was intentional, and (4) knowledge or reasonable grounds to know that removal would likely induce, enable, facilitate, or conceal an infringement. 17 U.S.C. § 1202(b)(1); *Mango*, 970 F. 3d at 171. Copyright registrations are not required, as OpenAI conceded in its Answer. *See Gattoni v. Tibi*,

LLC, 254 F. Supp. 3d 659, 663-64 (S.D.N.Y. 2017); OpenAI Answer to CIR Second Amend. Compl. ¶ 9, Dkt. 1652.

Plaintiffs seek summary judgment on the first three elements, which will streamline the issues for trial. Fed. R. Civ. P. 56(g).

A. Plaintiffs conveyed CMI in connection with copies of their Asserted Works.

CMI includes the title, notice of copyright, author, publisher, and terms of use. 17 U.S.C. § 1202(c)(1)-(6). Links to CMI qualify. 17 U.S.C. § 1202(c)(7).

Courts interpret “conveyed in connection with” “in its broadest sense.” *Janik v. SMG Media, Inc.*, 2018 WL 345111, at *12 (S.D.N.Y. Jan. 10, 2018) (quoting S. Rep. 105-190 at 35 (1998)). Thus, CMI need not be placed in close “[p]hysical proximity” to the work. *Pierson v. Infinity Music & Ent., Inc.*, 300 F. Supp. 3d 390, 396 (D. Conn. 2018). CMI at the top or the bottom of a webpage, including a copyright notice and a link to terms of use, is conveyed in connection with a work when it is “easily accessible in conjunction with” the work. *Id.* at 391-92, 396. Plaintiffs published their CMI either directly above the article (for author and title information) or in an easily accessible website footer (for copyright notice and terms of use information), thus conveying CMI “in connection with” copies of their Asserted Works. SF1257-1284.

CMI must be conveyed in connection with “copies” of the work. 17 U.S.C. § 1202(c). The Copyright Act defines “copy” in 17 U.S.C. § 101. This definition is satisfied by substantial similarity. *Knitwaves, Inc. v. Lollytogs Ltd. (Inc.)*, 71 F.3d 996, 1002 (2d Cir. 1995). And the standard definition applies to “copies” in the definition of CMI. *See ADR Int’l Ltd. v. Inst. for Supply Mgmt. Inc.*, 667 F. Supp. 3d 411, 427 (S.D. Tex. 2023). That definition is satisfied here because the copies from which OpenAI removed Plaintiffs’ CMI are either [REDACTED]

[REDACTED] or (2) [REDACTED]

[REDACTED] SF809, 814-819, 823, 1190-1191, 1263-1264, 1273-1275, 1281-1284, 1300-1301, 1360. Because the copies are identical or nearly identical, they are substantially similar.

B. OpenAI removed Plaintiffs' CMI without authority.

OpenAI removed CMI from Plaintiffs' Asserted Works without Plaintiffs' permission. SF1285-1299. Common Crawl scraped exact copies of Plaintiffs' websites with their CMI, and OpenAI downloaded [REDACTED] from Common Crawl. SF814-819, 1263-1264, 1273-1275, 1281-83. When OpenAI obtained those copies, they contained Plaintiffs' CMI. SF1257-1284. But CMI was [REDACTED] SF1361-1421. [REDACTED] [REDACTED]. SF1381-1407, 1417-1420. For example, The Intercept's copyright notice is absent from [REDACTED] in OpenAI's training sets, and Mother Jones's is absent in [REDACTED] percent of cases. SF1417-1418. Similarly, The Times's copyright notice is absent in [REDACTED] percent of cases; and DNP's copyright notice is absent in [REDACTED] percent of cases for five of the plaintiffs, and at least [REDACTED] percent for the remainder. SF1381-1389. ZD's copyright notice was stripped [REDACTED] percent of the time. SF1395, 1397-1398.

In creating training-set copies of the Asserted Works, OpenAI [REDACTED] [REDACTED] SF1307-1308, 1333-1344. [REDACTED] [REDACTED] SF1310-1315, 1317-1322, 1325-1332. For instance, in Asserted Works about copyright-related issues, [REDACTED] [REDACTED] SF1330-1331. OpenAI employees described the process as [REDACTED]

[REDACTED]. SF1304, 1347-1350, 1354, 1357. OpenAI's expert agreed. SF1304, 1349-1351.

OpenAI removed Plaintiffs' CMI at a massive scale:

Plaintiff	Number of Copies of Asserted Works from which OpenAI's Removed at Least One Piece of CMI
The Times	[REDACTED] (SF1363)
DNP	[REDACTED] (SF1365-1380)
The Intercept	[REDACTED] (pieces of CMI) (copies) (SF1412-1413)
CIR	[REDACTED] (pieces of CMI) (copies) (SF1415-16)
ZD	[REDACTED] (SF1395)

C. OpenAI's removal was intentional.

Intent exists when a defendant intentionally uses computer programs that are designed to remove information that qualifies as CMI. *N.Y. Times*, 777 F. Supp. 3d at 315. There is no dispute that OpenAI did exactly that. That is among the very purposes of its [REDACTED], as confirmed by OpenAI's employees and experts, who stated in documents and testimony that OpenAI

[REDACTED]. SF1303-1304, 1346-1359. OpenAI employees admitted that they [REDACTED]

[REDACTED] SF1357-1358. They also admitted that [REDACTED]

[REDACTED] SF1359. And Dr. Callison-Burch testified that [REDACTED]
[REDACTED] SF1355.

VI. Plaintiffs are entitled to separate awards of statutory damages for infringement of each article.

A court may award statutory damages "for all infringements involved in the action, with respect to any one work[.]" 17 U.S.C. § 504(c)(1). As discussed above, *see* Section III.A, each of

The Times's, DNP's, and CIR's Asserted Works and some ZD works were published in a newspaper or periodical issue for which they secured a timely copyright registration, and registration of the issue covers the individual articles. *Morris*, 259 F.3d at 68; U.S. Copyright Office, Circ. 62A at 2; 17 U.S.C. § 412.

In the Second Circuit, whether statutory damages are awarded for each individual work (article), or only on the collective work (issue), depends on whether, as of the date of infringement, the individual work was made available separately from the collective work. That is,

[W]hen a copyright holder or publisher issues material *on an independent basis*, the law permits a statutory damages award for each individual work.... In other words, our focus is on whether the plaintiff—the copyright holder—issued its works separately, or together as a unit....For these reasons, we conclude that the District Court properly allowed separate statutory damages awards for songs that the plaintiffs issued as singles, even if those songs were also made available on albums.

EMI Christian Music Grp., Inc. v. MP3tunes, LLC, 844 F.3d 79, 101 (2d Cir. 2016) (emphasis added). The Court approved the jury instruction that, “so long as the plaintiffs proved that ‘on the date of infringement, each individual [song] at issue in this case was available for purchase as a single [work],’ it could issue multiple statutory damages awards for singles that appeared on the same album.” *Id.*

The Asserted Works (articles) are produced and available for consumption as single works separate and apart from any issue in which they appear. At the time of first infringement in this case (SF139, 142-143), each article was available online for reading and downloading separate from the “issue” in which the article also appeared. SF441, 460-462. OpenAI’s expert admitted that by [REDACTED]

[REDACTED]

[REDACTED] SF442. The uncontested record

establishes that the Court should enter summary judgment that statutory damages, if awarded, should be on a per-article basis.

VII. The Court should grant summary judgment on Defendants’ remaining affirmative defenses.

In addition to fair use, Defendants have raised a litany of other affirmative defenses. *See* CMC Statement Exhibit A – Case Information, Dkt. 47-1. Under *Celotex*, Plaintiffs contend that Defendants have not provided evidence establishing these defenses. *See Celotex Corp. v. Catrett*, 477 U.S. 317, 325 (1986); *Tardif v. City of New York*, 991 F.3d 394, 403 (2d. Cir. 2021). The Court should therefore grant summary judgment to Plaintiffs on all of them.

VIII. Conclusion

The Court should grant summary judgment for Plaintiffs on (1) their *prima facie* claims of infringement by (a) acquiring, (b) training on, (c) grounding on (by Microsoft), (d) outputting (by Microsoft), and (e) [REDACTED] their asserted works; (2) Defendants’ affirmative defense of fair use for any of these infringements; (3) the elements of conveyance of CMI and its intentional removal without authority under the DMCA; (4) Plaintiffs’ entitlement to separate awards of statutory damages for each infringed article published in their newspapers and periodicals; and (5) Defendants’ remaining affirmative defenses.

Dated: September 4, 2026

/s/ Ian Crosby

Ian Crosby (*pro hac vice*)
Katherine M. Peaslee (*pro hac vice*)
Susman Godfrey L.L.P.
401 Union Street, Suite 3000
Seattle, WA 98101
Telephone: (206) 516-3880
Facsimile: (206) 516-3883
icrosby@susmangodfrey.com
kpeaslee@susmangodfrey.com

Davida Brook (*pro hac vice*)
Emily K. Cronin (*pro hac vice*)
Adnan Muttalib (*pro hac vice*)
Susman Godfrey L.L.P.
1900 Ave of the Stars, Suite 1400
Los Angeles, CA 90067
Telephone: (310) 789-3100
Facsimile: (310) 789-3150
dbrook@susmangodfrey.com
ecronin@susmangodfrey.com
amuttalib@susmangodfrey.com

Elisha Barron (5036850)
Zachary B. Savage (ZS2668)
Alexander Frawley (5564539)
Eudokia Spanos (5021381)
Susman Godfrey L.L.P.
One Manhattan West
New York, NY 10001
Telephone: (212) 336-8330
Facsimile: (212) 336-8340
ebarron@susmangodfrey.com
zsavage@susmangodfrey.com
afrawley@susmangodfrey.com
espanos@susmangodfrey.com

Steven Lieberman (SL8687)
Jennifer B. Maisel (5096995)
Kristen J. Logan (*pro hac vice*)
Rothwell, Figg, Ernst & Manbeck, P.C.
901 New York Avenue, N.W., Suite 900 East
Washington, DC 20001
Telephone: (202) 783-6040
Facsimile: (202) 783 6031
slieberman@rothwellfigg.com
jmaisel@rothwellfigg.com
klogan@rothwellfigg.com

Attorneys for Plaintiff
The New York Times Company

/s/ Steven Lieberman

Steven Lieberman (SL8687)

Jennifer B. Maisel (5096995)

Robert Parker (*admitted pro hac vice*)

Jenny L. Colgate (*admitted pro hac vice*)

Alexandra Hughes (*admitted pro hac vice*)

Mark Rawls (*admitted pro hac vice*)

Michael H. Jones (*admitted pro hac vice*)

Kristen J. Logan (*admitted pro hac vice*)

Bryan B. Thompson (6004147)

Mary L. Mullins (*pending pro hac vice*)

ROTHWELL, FIGG, ERNST & MANBECK,
P.C.

901 New York Avenue, N.W., Suite 900 East

Washington, DC 20001

Telephone: (202) 783-6040

Facsimile: (202) 783-6031

slieberman@rothwellfigg.com

jmaisel@rothwellfigg.com

rparker@rothwellfigg.com

jcolgate@rothwellfigg.com

ahughes@rothwellfigg.com

mrawls@rothwellfigg.com

mjjones@rothwellfigg.com

klogan@rothwellfigg.com

bthompson@rothwellfigg.com

mmullins@rothwellfigg.com

Attorneys for Daily News Plaintiffs

/s/ Matthew Topic

Jon Loevy (*pro hac vice*)
Michael Kanovitz (*pro hac vice*)
Matthew Topic (*pro hac vice*)
Thomas Kayes (*pro hac vice*)
Steven Art (*pro hac vice*)
Kyle Wallenberg (*pro hac vice*)
Shelley Geiszler (*pro hac vice*)
LOEVY & LOEVY
311 North Aberdeen, 3rd Floor
Chicago, IL 60607
312-243-5900 (p)
312-243-5902 (f)
jon@loevy.com
mike@loevy.com
matt@loevy.com
kayes@loevy.com
steve@loevy.com
wallenberg@loevy.com
geiszler@loevy.com

Stephen Stich Match (No. 5567854)
Lauren Sonnenberg (*pro hac vice*)
LOEVY & LOEVY
222 Sutter Street, Suite 600A
San Francisco, CA 94108
(312) 243-5900
match@loevy.com
sonnenberg@loevy.com

*Attorneys for Plaintiffs The Center for
Investigative Reporting, Inc. and The
Intercept Media, Inc.*

/s/ Lacy H. Koonce, III

Lacy H. ("Lance") Koonce, III
Matthew A. Leish
Brendan Kehoe
Lucy Hoffman
Maya Katalan
Klaris Law PLLC
161 Water Street
New York, NY 10038
Telephone: (646) 779-4882
lance.koonce@klaris.com

matthew.leish@klarislaw.com
brendan.kehoe@klarislaw.com
lucy.hoffman@klarislaw.com
maya.katalan@klarislaw.com

Guy Ruttenberg (*pro hac vice*)
Ruttenberg IP Law, A Professional
Corporation
1801 Century Park East
Suite 1920
Los Angeles, CA 90067
Telephone: (310) 627-2270
ZD-AICopyright@ruttenbergiplaw.com

John R. Cuti
Cuti Frisch PLLC
233 Broadway, Suite 1720
New York, NY 10279

*Attorneys for Plaintiffs Ziff Davis, Inc., Ziff
Davis, LLC, IGN Entertainment, Inc.,
Everyday Health Media, LLC, Mashable,
Inc., and CNET Media, Inc.*